

66250

From: Chan, Christina  
Sent: Wednesday, May 08, 2002 11:04 AM  
To: Davis, Minh-Tam; STIC-Biotech/ChemLib  
Subject: RE: Rush search request for 09/802520

Please rush. Thanks Chris

-----Original Message-----

Fr m: Davis, Minh-Tam  
Sent: Wednesday, May 08, 2002 10:53 AM  
To: Chan, Christina  
Subject: Rush search request for 09/802520

Please in commercial data base and in issued patent files

1) SEQ ID NO:1 and 2.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12  
305-2008

CRFE

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

## FOR OFFICIAL USE ONLY

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/8/02  
Date Completed: 5/2/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: OS  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      May 9, 2002, 00:03:01 ; Search time 31.89 Seconds
              (without alignments)
              375.307 Million cell updates/sec
```

Title: US-09-802-520-1  
 Perfect score: 2534  
 Sequence: 1 MESISMGSPKSLSETCLPN.....LEEGIGTIPHVSPEPTYVM 490

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1: /cgn2_6/prodata/2/1aa/5A.COMB.pcp.*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pcp.*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pcp.*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pcp.*
5: /cgn2_6/prodata/2/1aa/PCRTUS.COMB.pcp.*
6: /cgn2_6/prodata/2/1aa/backfilltest.pcp.*

```

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

No.						Description					
Score		Query		Length		DB		ID			
1	901	35.6	173	4	US-09-333-873A-8	Sequence 8, Appl					
2	788	31.1	339	4	US-09-333-873A-2	Sequence 2, Appl					
3	679	26.8	141	3	US-09-083-521-1	Sequence 1, Appl					
4	144.5	5.7	227	4	US-09-655-270A-15	Sequence 15, App					
5	144.5	5.7	227	4	US-09-651-941-17	Sequence 17, App					
6	118.5	4.7	531	2	US-08-724-974A-3	Sequence 3, Appl					
7	112.5	4.4	365	2	US-08-724-974A-3	Sequence 3, Appl					
8	107	4.2	34	4	US-09-333-873A-20	Sequence 20, App					
9	106	4.2	695	3	US-08-487-886-2	Sequence 2, Appl					
10	106	4.2	695	3	US-08-482-855-2	Sequence 2, Appl					
11	101	4.0	476	4	US-09-316-083-3	Sequence 3, Appl					
12	97.5	3.8	390	3	US-08-460-576-2	Sequence 2, Appl					
13	93	3.7	692	4	US-07-757-342D-6	Sequence 6, Appl					
14	89.5	3.5	495	1	US-07-841-997A-2	Sequence 2, Appl					
15	89.5	3.5	495	1	US-08-290-301-2	Sequence 2, Appl					
16	89.5	3.5	495	2	US-08-588-983-2	Sequence 2, Appl					
17	89.5	3.5	495	2	US-08-568-976-2	Sequence 2, Appl					
18	89.5	3.5	861	3	US-09-032-875-2	Sequence 2, Appl					
19	89.5	3.5	968	3	US-08-651-999A-7	Sequence 7, Appl					
20	89.5	3.5	968	4	US-09-385-752-7	Sequence 7, Appl					
21	89	3.5	467	4	US-08-805-118-3	Sequence 3, Appl					
22	89	3.5	467	4	US-09-331-958-3	Sequence 3, Appl					
23	89	3.5	1582	2	US-08-404-531B-9	Sequence 9, Appl					
24	89	3.5	1582	3	US-08-476-900A-9	Sequence 9, Appl					
25	89	3.5	1582	3	US-08-468-546A-9	Sequence 9, Appl					
26	89	3.5	1582	4	US-08-726-320-5	Sequence 5, Appl					
27	89	3.5	1582	4	US-09-208-716-5	Sequence 5, Appl					

28	88.5	3.5	591	1	US-08-484-840-2	Sequence 2, Appl1
29	88.5	3.5	591	1	US-08-483-094-2	Sequence 2, Appl1
30	88.5	3.5	699	1	US-08-348-006B-7	Sequence 7, Appl1
31	88.5	3.5	699	2	US-08-800-828A-7	Sequence 7, Appl1
32	88.5	3.5	699	4	US-09-158-657-7	Sequence 7, Appl1
33	88	3.5	920	3	US-08-930-996A-8	Sequence 8, Appl1
34	86	3.4	205	3	US-08-839-711-1	Sequence 1, Appl1
35	86	3.4	480	2	US-08-724-392A-9	Sequence 9, Appl1
36	85	3.4	337	1	US-08-153-844-46	Sequence 46, Appl
37	85	3.4	337	3	US-09-299-844A-46	Sequence 46, Appl
38	85	3.4	337	4	US-09-088-337B-46	Sequence 46, Appl
39	85	3.4	337	5	PCT-US93-11153-46	Sequence 46, Appl
40	85	3.4	352	1	US-08-202-055-3	Sequence 3, Appl1
41	85	3.4	352	1	US-08-076-093A-4	Sequence 4, Appl1
42	85	3.4	352	1	US-08-701-265-A	Sequence 4, Appl1
43	85	3.4	352	2	US-08-284-586-4	Sequence 4, Appl1
44	85	3.4	352	2	US-08-805-478-4	Sequence 4, Appl1
45	85	3.4	352	2	US-08-802-627A-4	Sequence 4, Appl1

## ALIGNMENTS

```

RESULT      1
US-09-323-873A-8
: Sequence: 8, Application US/09323873A
: Patient No. 6329503
: GENERAL INFORMATION:
: APPLICANT: Daniel E. Afar
: APPLICANT: Rene S. Hubert
: APPLICANT: Kahan Leong
: APPLICANT: Arthur B. Raitano
: APPLICANT: Douglas C. Saffran
: APPLICANT: Steve Chappell Mitchell
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: FILE REFERENCE: 129.160502
: CURRENT APPLICATION NUMBER: US/09/323.873A
: PRIOR APPLICATION NUMBER: 60/087,520
: PRIOR FILING DATE: 1998-06-01
: PRIOR APPLICATION NUMBER: 60/091,183
: PRIOR FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 173
: TYPE: PRT
: ORGANISM: Homo Sapiens
US-09-323-873A-8

Query Match      35.6%; Score 901; DB 4; Length 173;
Best Local Similarity 100.0%; Prcd. No. 1.8e-85;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      246 DEYKPIETIVNKKTPIVATITLSTIVYLAGLLAAAYQVYGCYRRPPLMTLWLOCRKOL 305
DB      1 DFKPIPIETVNTPIVATITLSTIVYLAGLLAAAYQVYIGYRRPPLMTLWLOCRKOL 60

QY      306 GILSEFFAMVHVAAYSLCPMRRSERYELFLNMAVQOVHANINENSNNNEEYWRILEMTISFCI 365
DB      61 GILSEFFAMVHVAAYSLCPMRRSERYELFLNMAVQOVHANINENSNNNEEYWRILEMTISFCI 120

QY      366 MSIGLISLAVTSPISVSNALNMRREFSTQSTLGYVALLISFHVLIYGMKRA 418
DB      121 MSIGLISLAVTSPISVSNALNMRREFSTQSTLGYVALLISFHVLIYGMKRA 173

RESULT      2
US-09-323-873A-2
: Sequence 2, Application US/09323873A
: Patient No. 6329503

```

**THIS PAGE BLANK (USPTO)**





**THIS PAGE BLANK (USPTO)**

AUTHORS Gattung, S. and Maggill, L.  
 TITLE The sequence of H. sapiens BAC clone RG016J04  
 JOURNAL Unpublished (1997)  
 REFERENCE 2 (bases 1 to 156214)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAY-1997)  
 COMMENT Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63108, USA  
 http://genome.wustl.edu/gsc  
 e-mail: sapientewatson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov)

SOURCE INFORMATION:  
 This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
 VECTOR: pBELO  
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
 Actual start of this clone is at base position 1 of H\_RG016J04; actual end is at 156214 of H\_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS893 (NID:g454733).

# FEATURES

source  
 Location/Qualifiers  
 1. 156214  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q21"  
 /clone="RG016J04"  
 /clone\_lib="CTB-978SK-B"  
 15. 40  
 /rpt\_family="L1"  
 repeat\_region  
 326. 621  
 /rpt\_family="ALU"  
 complement(977. 1499)  
 repeat\_region  
 /rpt\_family="L1"  
 3398. 3421  
 /rpt\_family="L1"  
 repeat\_region  
 5319. 5345  
 /rpt\_family="L1"  
 repeat\_region  
 complement(8071. 8347)  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(9406. 9975)  
 /rpt\_family="L1"  
 repeat\_region  
 complement(10000. 11285)  
 /rpt\_family="L1"  
 repeat\_region  
 complement(11315. 11984)

repeat\_region  
 /rpt\_family="L1"  
 11665. 11687  
 complement(12010. 12299)  
 repeat\_region  
 /rpt\_family="ALU"  
 complement(12301. 13893)  
 repeat\_region  
 /rpt\_family="L1"  
 13897. 14096  
 /rpt\_family="L1"  
 repeat\_region  
 complement(14777. 14838)  
 /rpt\_family="L1"  
 15715. 15767  
 /note="similar to human EST T02878 (NID:g314119)"  
 19436. 19497  
 /note="similar to human EST AA123941 (NID:g1682616)"  
 mq22709.r1  
 complement(25113. 25148)  
 /rpt\_family="L1"  
 complement(25561. 25578)  
 /rpt\_family="L1"  
 complement(26832. 27124)  
 /rpt\_family="ALU"  
 27125. 27383  
 /note="similar to human EST AA151807 (NID:g1720502)"  
 213902.r1  
 27125. 27544  
 /note="similar to human EST AA151796 (NID:g1720491)"  
 213902.r1  
 complement(27178. 27600)  
 /note="similar to human EST AA149579 (NID:g1720380)"  
 213902.r1  
 complement(27441. 27890)  
 /note="similar to human EST N52554 (NID:g1193720)"  
 yv36a11.s1  
 27686. 27856  
 /note="similar to human EST W32120 (NID:g1313113)"  
 zb97c08.r1  
 27686. 27838  
 /note="similar to human EST W32085 (NID:g1313105)"  
 zb97b08.r1  
 complement(27733. 28143)  
 /note="similar to human EST N59831 (NID:g1203721)"  
 yz77a10.s1  
 complement(27838. 28143)  
 /note="similar to human EST W31561 (NID:g1312680)"  
 zb97b08.s1  
 complement(27887. 28143)  
 /note="similar to human EST N59830 (NID:g1203720)"  
 yz77a09.s1  
 complement(27997. 28144)  
 /note="similar to human EST W31628 (NID:g1312688)"  
 zb97c08.s1  
 28666. 28972  
 /note="similar to human EST R09339 (NID:g761262)"  
 yf26d05.r1  
 complement(29587. 29853)  
 /note="similar to human EST R09227 (NID:g761150)"  
 yf26d05.s1  
 30146. 30314  
 /note="similar to human EST N78042 (NID:g1240743)"  
 yv71e06.r1  
 complement(31170. 31350)  
 /note="similar to human EST N58451 (NID:g1202341)"  
 yv71e06.s1  
 31970. 32006  
 /rpt\_family="L1"  
 repeat\_region  
 32653. 32563  
 /rpt\_family="ALU"  
 complement(34737. 34778)  
 /rpt\_family="L1"  
 36558. 36709  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(37630. 37740)

**THIS PAGE BLANK (USPTO)**

```

repeat_region      /rpt_family="PTR"
39839      .40125
repeat_region      /rpt_family="ALU"
40126      .40157
repeat_region      /rpt_family="L1"
complement(41340) .41706
repeat_region      /rpt_family="ALU"
complement(41904) .41933
repeat_region      /rpt_family="L1"
42348      .42409
repeat_region      /rpt_family="ALU"
complement(42942) .42982
repeat_region      /rpt_family="L1"
complement(44380) .44423
repeat_region      /rpt_family="L1"
44659      .44954
repeat_region      /rpt_family="ALU"
45422      .45448
repeat_region      /rpt_family="L1"
complement(45615) .45669
repeat_region      /rpt_family="L1"
complement(45672) .45963
repeat_region      /rpt_family="ALU"
complement(45964) .46597
repeat_region      /rpt_family="L1"
46100      .46516
repeat_region      /rpt_family="L1"
48370      .48785
repeat_region      /rpt_family="L1"
complement(49577) .50216
repeat_region      /rpt_family="L1"
complement(50540) .50791
repeat_region      /rpt_family="THP"
complement(50820) .50985
repeat_region      /rpt_family="THP"
complement(51124) .51156
repeat_region      /rpt_family="L1"
51405      .51691
repeat_region      /rpt_family="ALU"
complement(51721) .51764
repeat_region      /rpt_family="L1"
complement(51765) .52058
repeat_region      /rpt_family="ALU"
complement(52252) .52467
Query Match      28.2% Score 533.4 DB 9 Length 156214;
Best Local Similarity 99.8% Pred. No. 2e-124;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 804 caggttatatagagaacaataatccaagcgcagacaagattgaactggccgc 863
Db 19241 CAGGTTTATATATGAGCAACAAATTCAACGCGACAAAGGTTATTTGAACCTGCCGC
OY 864 cagttgaattcattccattgagcttgagctcattatcatcagcagagagattgaaat 923
Db 19301 CAGTTGAATTCATTTCCCATTTGAGGATCCCTTATCATCAGCCAGAGAGATTGAAAT
OY 924 ttaccctcagactcttactctctgagagagggcaggtgtgagctataaagctggcc 983
Db 19361 TTACCCCTAGAGACTCTTACTCTGTGAGAGGGCCACTGGGTAGCTATAGCTTGGCC
OY 984 acatttttcttattctctctgtcagagatgtgattcatcatcatagctagaaccaa 1043
Db 19421 ACAATTTTCTTCTTATATCTTTGTCAGAGATGTGATTCATCATATGCTAAGAAACCA
OY 1044 cagaagtactttacaaaattcctatagagattgtgaaataaaaacttaactagtgc 1103
Db 19481 CAGAGTGACTTTTACAAAATTCCTATAGAGATTGTGAATAAAACCTTACCTTAGTGGC
OY 1104 attacttgctctccctagataactcagaggtctcttgagagctgctataaacttat 1163
Db 19541 ATTACTTGCTCTCCCTAGTATACCTCGCAGAGTCTTCTGCGAGCTGCTTATCACTTAT
19600

```

```

OY 1164 taggcacccaagtataagagattccacactgtgttggaacctgtttacagttagaana 1223
Db 19601 TAGGCACCAAGTATAGAGATTTCACACTTGTGTGGAACCTGTTGACAGTGAAGAAA
OY 1224 cagcttgattactaagttttcttctgcgtatggtccatgltgctacagaccttgctta 1283
Db 19661 CAGCTTGATTACTAAGTTTTTCTTCGTATGCTGATGCTCATTTCCCTACACCTCTGCTTA
OY 1284 ccgattagagaagtcacagagatatgttttcttaacatggtctatacagagcttc 1338
Db 19721 CCGATGAGAGAGTCACAGAGATTTTGTTCACACATGCGTATACAGCAGGTAC 19775

RESULT 3
AC104475/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 162928)
Ahler, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, R., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Larcic, P., Lee-Tsin, S.-O.,
Legsaki, R., Maduro, O.L., Maduro, V.B., Mastello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantipoo, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Whiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 162928)
Green, E.D.
Direct Submission
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nih.gov
----- Project Information
Center project name: ces
Center clone name: 120K11

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161422 bases at least Q30
Insert size: 143000; agarose-fp
Insert size: 162428; sum-of-ctnigs
Quality coverage: 10.76x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-ctnigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 13922: contig of 13922 bp in length  
 \* 13923 14022: gap of unknown length  
 \* 14023

**THIS PAGE BLANK (USPTO)**

Query Match	44.38;	Score 229;	DB 6;	Length 2453
Best Local Similarity	100.0%;	Pred. No. 2.3e-128;		
Matches 229;	Conservative			

RESULT	2
HSAC002064	
LOCUS	HSAC002064
DEFINITION	Human BAC clone RGO1604 from 7q21, complete sequence.
ACCESSION	AC002064
VERSION	AC002064.1
	156214 bp DNA linear PRI 09-MAY-1997

St. Louis, MO 63108, USA  
<http://genome.wustl.edu/gsc>  
 e-mail: [saplen@watson.wustl.edu](mailto:saplen@watson.wustl.edu)

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.ngri.nih.gov/DIR/CTM/CHR7> or send an e-mail to [greene@ngri1.nih.gov](mailto:greene@ngri1.nih.gov)

**Library:** A library of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 9785K. For references see: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

**VECTOR:** pBelo

**Selection:** chloramphenicol

Actual end is at 156214 of H\_RG016J04. The orientation of this clone is unknown.

This clone contains STS **SWSS2784** (NID:q1113580) and **SWSS8933** (NID:q454733).  
Location: 70,111,161

```

source
location/Qualifiers
1..156214
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21"
/clone="RG016J04"
/location.lib="CITB-978SK-B"
15..40
/rpt._region="L1"
326..621
/rpt._family="ALU"
complement(977..1499)
/rpt._family="L1"
3398..3421
/rpt._family="L1"
5319..5345
/rpt._family="L1"
complement(12010..12299)
/rpt._family="L1"
complement(11315..11984)
/rpt._family="L1"
11666..11687
/rpt._family="L1"
complement(12010..12299)
/rpt._family="ALU"
complement(12301..13893)
/rpt._family="L1"
13897..14096
/rpt._family="L1"
complement(14777..14838)
/rpt._family="L1"
15715..15767
/note="similar to human EST T02878 (NTD:g314119)."
19436..19497
/note="similar to human EST AA123941 (NTD:g1682616
mq22109..r1"
complement(25113..25148)
/rpt._family="L1"
complement(25561..25578)
/rpt._family="L1"
complement(26832..27124)
/rpt._family="ALU"
27125..27383

```

**THIS PAGE BLANK (USPTO)**



misc\_feature /note="similar to human EST AA151807 (NID:g1720502)  
213902.r1"  
27125. .27544  
/note="similar to human EST AA151796 (NID:g1720491)  
213902.r1"  
complement(27178. .27600)  
/note="similar to human EST AA149579 (NID:g1720380)  
213902.r1"  
complement(27441. .27890)  
/note="similar to human EST N52554 (NID:g1193720)  
27686. .27856  
/note="similar to human EST W32120 (NID:g1313113)  
2b97c08.r1"  
27686. .27838  
/note="similar to human EST W32085 (NID:g1313105)  
2b97b08.r1"  
complement(27733. .28143)  
/note="similar to human EST N59831 (NID:g1203721)  
y27a10.r1"  
complement(27838. .28143)  
/note="similar to human EST W31561 (NID:g1312680)  
2b97b08.r1"  
complement(27887. .28143)  
/note="similar to human EST N59830 (NID:g1203720)  
y27a09.r1"  
complement(27997. .28144)  
/note="similar to human EST W31628 (NID:g1312688)  
2b97c08.r1"  
28666. .28972  
/note="similar to human EST R09339 (NID:g761262)  
y126d05.r1"  
complement(29587. .29853)  
/note="similar to human EST R09227 (NID:g761150)  
y126d05.r1"  
30146. .30314  
/note="similar to human EST N78042 (NID:g1240743)  
y17e06.r1"  
complement(31170. .31350)  
/note="similar to human EST N58451 (NID:g1202341)  
y17e06.r1"  
31970. .32006  
/rpt\_family="L1"  
repeat\_region /rpt\_family="L1"  
32263. .32563  
/rpt\_family="ALU"  
complement(34737. .34778)  
/rpt\_family="L1"  
36558. .36709  
/rpt\_family="ALU"  
complement(37630. .37740)  
/rpt\_family="PTR"  
39839. .40125  
/rpt\_family="ALU"  
40126. .40157  
/rpt\_family="L1"  
repeat\_region /rpt\_family="L1"  
complement(41340. .41706)  
/rpt\_family="ALU"  
complement(41904. .41933)  
/rpt\_family="L1"  
42348. .42409  
/rpt\_family="ALU"  
complement(42942. .42982)  
/rpt\_family="L1"  
complement(44380. .44423)  
/rpt\_family="L1"  
44659. .44954  
/rpt\_family="ALU"  
45422. .45448  
/rpt\_family="L1"  
complement(45615. .45669)  
/rpt\_family="L1"  
complement(45672. .45963)  
/rpt\_family="ALU"

repeat\_region complement(45964. .46597)  
/rpt\_family="L1"  
repeat\_region 46100. .46516  
/rpt\_family="L1"  
repeat\_region 48370. .48785  
/rpt\_family="L1"  
repeat\_region complement(49577. .50216)  
/rpt\_family="L1"  
repeat\_region complement(50540. .50791)  
/rpt\_family="HE"  
repeat\_region complement(51124. .51156)  
/rpt\_family="L1"  
repeat\_region 51405. .51691  
/rpt\_family="ALU"  
repeat\_region complement(51721. .51764)  
/rpt\_family="L1"  
repeat\_region complement(51765. .52058)  
/rpt\_family="ALU"  
repeat\_region complement(52252. .52467)  
Query Match 34.4% Score 178; DB 9; Length 156214;  
Best Local Similarity 99.6%; Pred. No. 7.5e-97;  
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 280 aggaatctctgtgtagcttggaaggtccgcatcatggaatcctcatgagga 339  
|||||  
Db 17321 AGGATATCTTGTCATCTTGGAAGTCCGATCATGGAATCATCTGATGAGGGA 17380  
|||||  
Oy 340 gccctaagagccttagaactgttcaactaatgacataaagtatcaagaatgcaa 399  
|||||  
Db 17381 GCCCTAAGAGCCTTAGTGAACCTTTTACCTAATGACATAATGATCAAGATGCA 17440  
|||||  
Oy 400 ggaaggtcactgtgtagtctgattggaatgagatttgcacaaaccttgaccattgac 459  
|||||  
Db 17441 GGAAGGTCACTGTAGTGTGATGGAAGTGGAGATTTGCCAAATCTTGACATTGAC 17500  
|||||  
Oy 460 ttatagatgcgcctatcatctgtgcacatgagaagatgagaatccatgaatt 508  
|||||  
Db 17501 TTATTAGATCGCGCATCATGTGCTCATGACATGACATGACATGACATGATT 17549  
|||||  
RESULT 3  
AC104475 162928 bp DNA linear HTG 12-DEC-2001  
AC104475/c  
LOCUS  
DEFINITION Pan troglodytes clone RP43-120k11, WORKING DRAFT SEQUENCE, 6  
unordered pieces.  
ACCESSION AC104475.1 GI:17530717  
VERSION AC104475.1  
KEYWORDS HTG: HTGS PHASE1; HTGS\_DRAFT.  
SOURCE chimpanzee.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
1 (bases 1 to 162928)  
REFERENCE  
AUTHORS Akter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,  
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masello,C., Mastrian,S.D.,  
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Stantiripop,S.,  
Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L.,  
Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and  
Green,E.D.  
NISC Comparative Sequencing Initiative  
JOURNAL NISC Comparative Sequencing Initiative  
REFERENCE Unpublished  
2 (bases 1 to 162928)  
AUTHORS Green,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717  
Govermont Circle, Gaithersburg, MD 20877, USA  
COMMENT ----- Genome Center

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 21:01:15 ; Search time 3502.82 Seconds  
(without alignments)  
11297.206 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagctgagtgccg.....gtcaattatcgtgggttga 1891

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vl:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vl:\*  
30: em.htg.hum:\*  
31: em.htg.inv:\*  
32: em.htg.other:\*  
33: em.htg.inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	1409.8	74.6	2453	6	AX155249	AX155249 Sequence
2	533.4	28.2	156214	9	HSAC002064	AC002064 Human BAC
3	528.6	28.0	162828	2	AC104475	AC104475 Pan trogl
4	516.2	27.3	172915	2	AC099742	AC099742 Papio cyn
5	425.8	22.5	444	6	AX106433	AX106433 Sequence
6	425.8	22.5	444	6	AX140724	AX140724 Sequence
7	425.8	22.5	444	6	AX200584	AX200584 Sequence
8	425.8	22.5	444	6	AX267240	AX267240 Sequence
9	415.2	22.0	1971	10	AY029586	AY029586 Mus muscu
10	411	21.7	205085	2	AC026813	AC026813 Mus muscu
11	411	21.7	227144	2	AC092404	AC092404 Mus muscu
12	406.6	21.5	2707	10	AF335281	AF335281 Rattus no
13	381	20.1	3844	6	AX211372	AX211372 Sequence
14	379.4	20.1	1845	9	AY029585	AY029585 Homo sapi
15	379.4	20.1	3912	9	AK001691	AK001691 Homo sapi
16	363	19.2	366	6	AX106434	AX106434 Sequence
17	363	19.2	366	6	AX140725	AX140725 Sequence
18	363	19.2	366	6	AX200585	AX200585 Sequence
19	363	19.2	366	6	AX267241	AX267241 Sequence
20	358.2	18.9	1858	6	AX083516	AX083516 Sequence
21	356.6	18.9	4429	6	AX155251	AX155251 Sequence
22	337.8	17.9	2462	2	AK026806	AK026806 Homo sapi
23	335.2	17.7	87401	2	AC021898	AC021898 Homo sapi
24	318.2	16.8	3155	10	BC006651	BC006651 Mus muscu
25	317.8	16.8	1763	10	AY029778	AY029778 Mus muscu
26	315	16.7	3080	10	MMU0319746	MMU0319746 Mus muscu
27	310.2	16.4	322	6	AX155253	AX155253 Sequence
28	306.8	16.2	328	6	AX106431	AX106431 Sequence
29	306.8	16.2	328	6	AX140722	AX140722 Sequence
30	306.8	16.2	328	6	AX200582	AX200582 Sequence
31	306.8	16.2	328	6	AX267238	AX267238 Sequence
32	301.6	15.9	1754	4	AF319659	AF319659 Sus scrofa
33	290.4	15.4	1193	6	AX155243	AX155243 Sequence
34	290.4	15.4	1195	6	AX201105	AX201105 Sequence
35	290.4	15.4	1195	6	AX267904	AX267904 Sequence
36	290.4	15.4	1195	9	AF186249	AF186249 Homo sapi
37	290.4	15.4	1249	9	BC011802	BC011802 Homo sapi
38	272.8	14.4	1175	10	AF297098	AF297098 Homo sapi
39	266.4	14.1	1191	10	AY029584	AY029584 Mus muscu
40	243.4	12.9	250	6	AX106432	AX106432 Sequence
41	243.4	12.9	250	6	AX140723	AX140723 Sequence
42	243.4	12.9	250	6	AX200583	AX200583 Sequence
43	243.4	12.9	250	6	AX267239	AX267239 Sequence
44	232.2	12.3	3627	6	AX155248	AX155248 Sequence
45	228.4	12.1	1203	6	AF3553	AF3553 Sequence 6

#### ALIGNMENTS

RESULT 1  
AX155249  
LOCUS AX155249 2453 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 7 from Patent WO0140276.  
ACCESSION AX155249  
VERSION AX155249.1 GI:14536713  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homilidae; Homo.

REFERENCE  
1 (bases 1 to 2453)  
Farrar, D.E., Hubert, R.S., Raitano, A.B., Saffran, D.C., Mitchell, S.C.,  
Serpentine transmembrane antigens expressed in human prostate  
cancers and uses thereof  
Patent: WO 0140276-A 7 07-JUN-2001;  
JOURNAL Urogenesys, Inc. (US)  
FEATURES  
source  
1..2453  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
355..1719



AUTHORS	Galling, S. and Maggi, L.
TITLE	The sequence of H. sapiens BAC clone RG016J04.
JOURNAL	Unpublished (1997)
REFERENCE	2 (bases 1 to 156214)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-1997)
COMMENT	Genome Sequencing Center

Department of Genetics, Washington University  
St. Louis, MO 63108, USA  
<http://genome.wustl.edu/gsc>  
e-mail: [saplens@watsn.wustl.edu](mailto:saplens@watsn.wustl.edu)

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7> or send an E-mail to [sgreen@nhgri.nih.gov](mailto:sgreen@nhgri.nih.gov)

**SOURCE INFORMATION:**

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., *Proc. Natl. Acad. Sci.* 89:8794-8797 (1992); Kim et al., *Genomics* 34:213-218 (1996).

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of H\_RG016J04.;  
actual end is at 156214 of H\_RG016J04. The orientation of this  
clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS8933 (NID:g454733).

## FEATURES

**Source**

```
repeat_region
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21"
/clone="RG016J04"
/cclone_id="CITB-978SK-B
15..40
/rpt_family="L1"
326..621
/rft_family="ALU"
complement(977..1499)
rpt_region
/rft_family="L1"
3398..3431
/rft_family="L1"
/rft_family="L1"
5319..5345
/rft_family="L1"
complement(8071..8347)
rpt_region
/rft_family="ALU"
complement(9406..9975)
rft_family="L1"
complement(10000..11285)
rft_family="L1"
complement(11315..11984)
rpt_region
```

repeat_region	/rpt_family="L1"	11666 .11687
repeat_region	/rpt_family="L1"	complement(12010 .12299)
repeat_region	/rpt_family="ALU"	complement(12301 .13893)
repeat_region	/rpt_family="L1"	13897 .14096
repeat_region	/rpt_family="L1"	complement(14777 .14838)
misc_feature	/rpt_family="L1"	15715 .15767
misc_feature	/note="similar to human EST T02878 (NID:g314119)"	19436 .19497
misc_feature	/note="similar to human EST AA123941 (NID:g1682616)	mq22f09.t1"
repeat_region	complement(25113 .25148)	/rpt_family="L1"
repeat_region	complement(25561 .25578)	/rpt_family="L1"
repeat_region	complement(26832 .27124)	/rpt_family="ALU"
misc_feature	/note="similar to human EST AA151807 (NID:g1720502)	z139g02.t1"
misc_feature	z139g02.t1"	27125 .27544
misc_feature	/note="similar to human EST AA151796 (NID:g1720491)	z139c02.t1"
misc_feature	complement(27178 .27600)	/note="similar to human EST AA149579 (NID:g1720380)
misc_feature	z139c02.s1"	complement(27441 .27890)
misc_feature	/note="similar to human EST N52554 (NID:g1193720)	yv36a11.s1"
misc_feature	/note="similar to human EST W21210 (NID:g1313113)	z2b97c08.t1"
misc_feature	z2b97c08.t1"	27866 .27856
misc_feature	/note="similar to human EST W32085 (NID:g1313105)	z2b97b08.s1"
misc_feature	complement(27733 .28143)	/note="similar to human EST N59831 (NID:g1203721)
misc_feature	yz77a10.s1"	complement(27838 .28143)
misc_feature	/note="similar to human EST W31561 (NID:g1312680)	zb97b08.s1"
misc_feature	complement(27887 .28143)	/note="similar to human EST N59830 (NID:g1203720)
misc_feature	yz77a09.s1"	complement(27997 .28144)
misc_feature	/note="similar to human EST W31628 (NID:g1312688)	zb97c08.s1"
misc_feature	28666 .28972	/note="similar to human EST R09339 (NID:g761262)
misc_feature	yz16d05.t1"	complement(29587 .29853)
misc_feature	/note="similar to human EST R09227 (NID:g761150)	yf6cd05.s1"
misc_feature	30146 .30314	/note="similar to human EST N78042 (NID:g1240743)
misc_feature	yy1ve06.t1"	complement(31170 .31350)
misc_feature	/note="similar to human EST N58451 (NID:g1202341)	yv1ve06.s1"
repeat_region	31970 .32006	/rpt_family="L1"
repeat_region	32263 .32563	/rpt_family="ALU"
repeat_region	complement(34737 .34778)	/rpt_family="L1"
repeat_region	36558 .36709	/rpt_family="ALU"
repeat_region	complement(37630 .37740)	

```

repeat_region      /rpl_family="PFR"
                    39839. .40125
                    /rpl_family="ALU"
repeat_region      40126. .40157
                    /rpl_family="L1"
repeat_region      complement(41340)
                    /rpl_family="ALU"
repeat_region      complement(41904)
                    /rpl_family="L1"
repeat_region      42348. .42409
                    /rpl_family="ALU"
repeat_region      complement(42942)
                    /rpl_family="L1"
repeat_region      complement(44380)
                    /rpl_family="L1"
repeat_region      44659. .44954
                    /rpl_family="ALU"
repeat_region      45422. .45448
                    /rpl_family="L1"
repeat_region      complement(45615)
                    /rpl_family="L1"
repeat_region      complement(45672)
                    /rpl_family="ALU"
repeat_region      complement(45964)
                    /rpl_family="L1"
repeat_region      46100. .46516
                    /rpl_family="L1"
repeat_region      48370. .48785
                    /rpl_family="L1"
repeat_region      complement(49577)
                    /rpl_family="L1"
repeat_region      complement(50540)
                    /rpl_family="L1"
repeat_region      complement(50820)
                    /rpl_family="L1"
repeat_region      complement(51124)
                    /rpl_family="L1"
repeat_region      51405. .51691
                    /rpl_family="ALU"
repeat_region      complement(51721)
                    /rpl_family="L1"
repeat_region      complement(51765)
                    /rpl_family="ALU"
repeat_region      complement(52252)
                    /rpl_family="L1"
repeat_region      52467
                    /rpl_family="L1"

Query Match      28.2%; Score 533.4; DB 9; Length 156214;
Best Local Similarity 99.8%; Pred. No. 2e-124;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1164 lacgaccacgaatagagattccaccttggttggaacccggtgtacagtgtaagaaa 1223
      TACGGACACCAAGTATAGCAATTTCCACTTGGTGGAAACCTGGTATACAGTGAGAAAA 19660
DB 19601 TACGGACACCAAGTATAGCAATTTCCACTTGGTGGAAACCTGGTATACAGTGAGAAAA 19660
QY 1224 cagcttgatctactaaattttctcttcgtcatagtgctatgttcctcctcctta 1283
      CAGCTTGATTTACTAAATTTTTCTTCTGCTATGATGCTCATTTCCATTCACCTTCCTTA 19720
DB 19661 CAGCTTGATTTACTAAATTTTTCTTCTGCTATGATGCTCATTTCCATTCACCTTCCTTA 19720
QY 1284 ccagtagaagatcagaagatatgttctcacaatgcttatcagcagctc 1338
      CCGATGAGAGGTCAGAGATATTTCTTCTCAGCATGCTTATACAGAGTAC 19775
DB 19721 CCGATGAGAGGTCAGAGATATTTCTTCTCAGCATGCTTATACAGAGTAC 19775

```

```

RESULT 3
AC104475/c
LOCUS
DEFINITION
Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
unordered pieces.
ACCESSION
AC104475
VERSION
AC104475.1 GI:17530717
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
chimpanzee.
ORGANISM
Pan troglodytes

```

```

REFERENCE
AUTHORS
Akher, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.M., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Lait, P., Lee, L.H., S.-Q.,
Legaspi, R., Maduro, Q., L., Maduro, V.B., Masello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantipop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgueon, C., Vogt, J.L.,
Walker, M.A., Metherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 162928)
Green, E.D.
Direct Submission
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nibrl.nih.gov
----- Project Information
Center project name: ccs
Center clone name: 120K11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161562 bases at least Q30
Insert size: 143000; agarose-fp
Insert size: 162428; sum-of-contigs
Quality coverage: 10.76x in Q20 bases; sum-of-contigs
Quality coverage: 9.47x in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 13922: contig of 13922 bp in length
* 13923 14022: gap of unknown length
* 14023 36248: contig of 22226 bp in length

```

```

*      36249      36348: gap of unknown length
*      36349      66192: contig of 29844 bp in length
*      66193      66292: gap of unknown length
*      66293      92168: contig of 25876 bp in length
*      92169      92268: gap of unknown length
*      92269      122036: contig of 29768 bp in length
*      122037      122136: gap of unknown length
*      122137      162928: contig of 40792 bp in length.
FEATURES
SOURCE
1. 162928
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-120K11"
/clone_1lb="RP43"
1. 13922
/note="assembly_fragment"
clone_end:17
vector_side:left"
14023. 36248
/note="assembly_fragment"
36349. 66192
/note="assembly_fragment"
66293. 92168
/note="assembly_fragment"
92269. 122036
/note="assembly_fragment"
122137. 162928
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

```

```

BASE COUNT      51528 a 29878 c 30533 g 50485 t 504 others
ORIGIN

```

```

Query Match      28.0%, Score 528.6; DB 2; Length 162928;
Best Local Similarity 99.3%; Pred. No. 3.4e-123;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      804      caggttatatagagaacaatatccaagcgagacaacaggttatgaacttccgcg 863
Db 121803 CAGGTTATATATGAGACACAAATATCAAGCCGCAACAGGTTATGTAACCTTCCCGC 121744

QY      864      cagtgtaattcattcccatgaacttggtggtatccatcatcagcagagagattgaat 923
Db 121743 CAGTTGATTTTCATTTCCCATGACTGGGATCCTTATCATCAGCCAGAGAGATTGAANAAT 121684

QY      924      ttaccctcagactcttactctctcgagagggccagatggtggtggtactataaagcttgcc 983
Db 121683 TTACCCCTACGACCTTACTCTCTGAGAGAGGGCCAGGTGTTAGCTATAAGCTTGGCC 121624

QY      984      acatttttttcccttattctcttcttgctcagagatggtatcatcatcatatgctagaaccaa 1043
Db 121623 ACATTTTTCCTTATTCCTTTCCTGTCAGAGATGATTCATCATATGCAAGAAACCA 121564

QY      1044      cagaatgactttacaacattccatagagattgtaataaacttaacttaacttaagcttgcc 1103
Db 121563 CAGAGTGACTTTTACAAAATTCCTATAGAGATTGGAATTAACCTTACTTATGTTGCC 121504

QY      1104      attacttgcctccctagatatactcgcagagctcttcgcagcagctgctatcaactttat 1163
Db 121503 ATTACTTGTCTCTCTCTAGTATACCTCGCAGAGTCTTCTGGCAGCTGCTTATCAACTTAT 121444

QY      1164      tacggcaccagaatataagagatttccacctggttggaacctggtttacagtgtagaaaa 1223
Db 121443 TACGGCACCAAGTATAGAGATTTCACCTTGGTGGAAACCTGTTACAGTGTGAAANA 121384

QY      1224      cagcttgatctactaagtttttcttcgcctatgctcagatgctcagcctcgcctta 1283
Db 121383 CAGCTTGATTTACTAAGTTTTTTCTTCCTGTGTCATGTTGCTTACAGCCTCTGCTTA 121324

QY      1284      ccgatgagaaggtcagagagatatattgttctcaacatgctcttaccagcaggttc 1338
Db 121323 CCGATGAGAAGGTCAAGAGATATTTGTTCTCAACATGGCTTATCACAGGTAC 121269

```

```

RESULT 4
AC099742
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nigrl.nih.gov
----- Project Information
Center project name: ccy
Center clone name: 167P22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 17615; sum-of-ctnigs
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-ctnigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
2438: contig of 2438 bp in length
2439
2538: gap of unknown length
2539
8133: contig of 5555 bp in length
8134
8233: gap of unknown length
8234
40378: contig of 32145 bp in length
40379
172915: contig of 132437 bp in length.
Location/Qualifiers
1. 172915
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone="RP41-167P22"
/clone_1lb="RP41"

```

```

FEATURES
SOURCE

```

```

misc-feature 1. 2438
               /note="assembly_fragment
               clone_end:SP6
               vector_side:right"
misc-feature 2539.8133
               /note="assembly_fragment"
               /db_xref="taxon:9606"
misc-feature 8234.40378
               /note="assembly_fragment
               clone_end:77
               vector_side:right"
misc-feature 40479.172915
               /note="assembly_fragment"
BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN

Query Match 27.3%; Score 516.2; DB 2; Length 172915;
Best Local Similarity 97.6%; Pred. No.5e-120;
Matches 524; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 280 aggaatcttggtgagcttggaagtgctgatacatgaatcatctatgagaa 339
    |||||
DB 83188 AGGATATTTTGGTGTGATCTTGGAAGTGTCTGTATCATGGAATCAATCTGTATGATGGAA 83247

QY 340 gccctaaagagccttagtgaactgtttacctaaaggcataaaagtgatcaaa 399
    |||||
DB 83248 GCCCTAAGAGCCCTTAACTGTAACCTTTTACCTAATGSCATAAATGATCAAGATGCAA 83307

QY 400 ggaagagcactagtgatgagtggaatggagatttgcacaaatccttgaccatgcac 459
    |||||
DB 83308 GGAAGACTACTGTAGTGTGATGGAGAGTGGAAATTTGCCAAATCTTGACGATTCGAC 83367

QY 460 ttattagatgcggctacatctgtgcataggaagtagaaatccctaagttgctctgaat 519
    |||||
DB 83368 TTATTAGATGTGGCTATCATGTGTCATAGGAAGTAGAATCTTAAGTTTGCCTTGAAAT 83427

QY 520 ttttccctcatgtgttagatgtcaatcatcaatgaagaagctctcacaaaaaataaa 579
    |||||
DB 83428 TTTTTCCTCATGTGTGATGTGCACCCATCATGAATGATCTCTCACAAAATAATATA 83487

QY 580 tattgtgctatacagaagaacattatacctcctgtggagccggagacatctgtg 639
    |||||
DB 83488 TATTGTCTTATACATFAGAGACATTTATCTCCTGTGGAGCTTAAGACATCTGCTGG 83547

QY 640 tgggtaaaaatccctgattgtatgtgagcaataaagatgaataaccagtaaccagaa 699
    |||||
DB 83548 TGGGTAAATCTGATTTGATGTGAGCAATATACATGAGATTAACCGATCCAGATCCA 83607

QY 700 atgcggaatatttgctcataatattcccaagattcttgattgtcaagaagattaatgtg 759
    |||||
DB 83608 ATGCTGAATATTTTGGCGCATATTATCCAGATTCTTTGATTGTCAAGAGATTTAATGTGG 83667

QY 760 tctcagcttgagcaactcagtttagagaccataagatccagccggaggtttataat 816
    |||||
DB 83668 TCTCAGCTTGGGCACCTTACGTTAGGACCTAAGATGCCAGCCGCGATATGATTT 83724

RESULT 5
AX106433 444 bp DNA linear PAT 30-APR-2001
LOCUS AX106433
DEFINITION Sequence 214 from Patent WO0125272.
ACCESSION AX106433
VERSION AX106433.1 GI:13922112
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 444)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
PATENT: WO 0125272-A 214 12-APR-2001;
JOURNAL

```

```

CORIXA CORPORATION (US)
FEATURES
    Location/Qualifiers
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
misc-feature 1.444
               /note="n = A,T,C or G"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;
Best Local Similarity 98.6%; Pred. No.1.1e-97;
Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 acccaaatccaatgctggaatatttggtcattatcccaagattccttgatgtcaaa 747
    |||||
DB 1 ACCCAGATCCAAATGCTGAATATTTGGCTTCATTTATCCAGATTTCTTATTTGTCAAG 60

QY 748 gattaaatgtgtcagcttgagcactlcaagtlaggaacctaaagatgccaagccgag 807
    |||||
DB 61 GATTTAATGTTGTCTCAGCTTGGGCACTTCAGTTAGACCTAAGATGCCAGCCGCGAG 120

QY 808 ttatataatgcagcaaatatcatcaagcgcgaaacaaaggttatagactggccgcagt 867
    |||||
DB 121 TTATATATGACACCAATATTTCAACGCGACACAGGTTATGAACTTGCCGCGAGT 180

QY 868 tgattatcattccatgacttgagatcctcatcatcaagcagagagatgaaatttac 927
    |||||
DB 181 TGAATTTCAATCCCATTTGACTTGGATCTTATCATCAGCCANAGAGATTTGAAATTTAC 240

QY 928 ccctacgactcttactctcttgagagagggcaagtggtgagtataagcttgcca- 966
    |||||
DB 241 CCTACGACTCTTACTCTCTGAGAGGGCCAGTGGGTAGCTATTAAGCTTGCGCCACAT 300

QY 987 ttttttcccttattccttctgtgcagagatgtatcatcatatgtctaaacccaag 1046
    |||||
DB 301 TTTTTCCTTATTTCTTGTTCAGAGATGCAATTCATCATATGCTANAAACCAACAG 360

QY 1047 agtgaactttacaatctctatagagatgtgtaataaacccttaactagttgccat 1106
    |||||
DB 361 AGTGACTTTACAAATTTCTTATAGANATTGTGAATTAACCTTACTATAGTTGCCATT 420

QY 1107 acttgctctccctagatactc 1130
    |||||
DB 421 ACTTGCTCTCCCTAATATATCTC 444

RESULT 6
AX140724 444 bp DNA linear PAT 31-MAY-2001
LOCUS AX140724
DEFINITION Sequence 214 from Patent WO0134802.
ACCESSION AX140724
VERSION AX140724.1 GI:14280835
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 444)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skeiky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
PATENT: WO 0134802-A 214 17-MAY-2001;
JOURNAL CORIXA CORPORATION (US)
FEATURES
    Location/Qualifiers
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

```



Query Match 22.5%; Score 425.8; DB 6; Length 444;  
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;  
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 acccagaatccatgctgtaataatggtctcattatccccaagattcttgatgtgcaag 747  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 ACCCAAAATCCAAATGCTGAATATTTGGCTTCATTTCCAGATCTTTGATGTCAAG 60

QY 748 gatttaatgtgtctcagcttgggcactcagttgaagacctaaagatgcccagcag 807  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTTAAGAGATCCAGCGGCAGG 120

QY 808 ttataatgcaagaacaatattcaagcgacacaaggttatggaacttgcccgccagt 867  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 TTTATATATGACAGCAACAATATTCAAGCGGACAGAGTTATTGAATGCGCCAGT 180

QY 868 tgaattcatcccatgacttggatcctatcatcagcagaagatgaaattac 927  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 TGAATTTCAATTCCTTACTTGCGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 928 cccatgaccttactcctctggaagggcagtggtgagctataagcttggccaca- 986  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 CCTAGACACTTTACTCTCTGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 987 ttttttcttattcctcttgcagagatgagatcatccatagtgaaccaag 1046  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 TTTTTCCTTATTCCTTGTGCAGAGATGCAATTCATATCCTANAAACCAACAG 360

QY 1047 agtgaactttacaataatcctataagagatgtaataaacttactaagttgccatt 1106  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 AGGACTTTTACAAAATTCCTATAGANATGTGAATTAACCTTACTATAGTTGCCATT 420

QY 1107 acttgctctccctagatatactc 1130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 7  
 AX200584 444 bp DNA linear PAT 29-AUG-2001  
 LOCUS AX200584  
 DEFINITION Sequence 214 from Patent W00151633.  
 ACCESSION AX200584  
 VERSION AX200584.1 GI:15390404  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 444)  
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
 Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Reltter,M.W.,  
 Stolk,J.A., Skeiky,Y.A., Wang,A. and Mesgher,M.J.  
 TITLE Compositions and methods for the therapy and diagnosis of prostate  
 cancer  
 JOURNAL Patent: WO 0151633-A 214 19-JUL-2001;  
 CORIXA CORPORATION (US)  
 FEATURES  
 source 1..444  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 120 a 100 c 78 g 143 t 3 others  
 ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;  
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;  
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 acccagaatccatgctgtaataatggtctcattatccccaagattcttgatgtgcaag 747  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 ACCCAAAATCCAAATGCTGAATATTTGGCTTCATTTCCAGATCTTTGATGTCAAG 60

QY 748 gatttaatgtgtctcagcttgggcactcagttgaagacctaaagatgcccagcag 807  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTTAAGAGATCCAGCGGCAGG 120

QY 808 ttataatgcaagaacaatattcaagcgacacaaggttatggaacttgcccgccagt 867  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 TTTATATATGACAGCAACAATATTCAAGCGGACAGAGTTATTGAATGCGCCAGT 180

QY 868 tgaattcatcccatgacttggatcctatcatcagcagaagatgaaattac 927  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 TGAATTTCAATTCCTTACTTGCGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 928 cccatgaccttactcctctggaagggcagtggtgagctataagcttggccaca- 986  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 CCTAGACACTTTACTCTCTGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 987 ttttttcttattcctcttgcagagatgagatcatccatagtgaaccaag 1046  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 TTTTTCCTTATTCCTTGTGCAGAGATGCAATTCATATCCTANAAACCAACAG 360

QY 1047 agtgaactttacaataatcctataagagatgtaataaacttactaagttgccatt 1106  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 AGGACTTTTACAAAATTCCTATAGANATGTGAATTAACCTTACTATAGTTGCCATT 420

QY 1107 acttgctctccctagatatactc 1130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 8  
 AX267240 444 bp DNA linear PAT 26-OCT-2001  
 LOCUS AX267240  
 DEFINITION Sequence 214 from Patent W00173032.  
 ACCESSION AX267240  
 VERSION AX267240.1 GI:16516018  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
 Kalos,M.D., Fanger,G.R., Reltter,M.W., Stolk,J.A., Day,C.H.,  
 Vedrick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.  
 and Henderson,R.A.  
 TITLE Compositions and methods for the therapy and diagnosis of prostate  
 cancer  
 JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;  
 CORIXA CORPORATION (US)  
 FEATURES  
 source 1..444  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 120 a 100 c 78 g 143 t 3 others  
 ORIGIN

QY 748 gatttaatgtgtctcagcttgggcactcagttgaagacctaaagatgcccagcag 807  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTTAAGAGATCCAGCGGCAGG 120

QY 808 ttataatgcaagaacaatattcaagcgacacaaggttatggaacttgcccgccagt 867  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 TTTATATATGACAGCAACAATATTCAAGCGGACAGAGTTATTGAATGCGCCAGT 180

QY 868 tgaattcatcccatgacttggatcctatcatcagcagaagatgaaattac 927  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 TGAATTTCAATTCCTTACTTGCGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 928 cccatgaccttactcctctggaagggcagtggtgagctataagcttggccaca- 986  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 CCTAGACACTTTACTCTCTGAGAGGCGCAGTGTGAGCTATAGCTTGGCCACAT 300

QY 987 ttttttcttattcctcttgcagagatgagatcatccatagtgaaccaag 1046  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 TTTTTCCTTATTCCTTGTGCAGAGATGCAATTCATATCCTANAAACCAACAG 360

QY 1047 agtgaactttacaataatcctataagagatgtaataaacttactaagttgccatt 1106  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 AGGACTTTTACAAAATTCCTATAGANATGTGAATTAACCTTACTATAGTTGCCATT 420

QY 1107 acttgctctccctagatatactc 1130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 8  
 AX267240 444 bp DNA linear PAT 26-OCT-2001  
 LOCUS AX267240  
 DEFINITION Sequence 214 from Patent W00173032.  
 ACCESSION AX267240  
 VERSION AX267240.1 GI:16516018  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
 Kalos,M.D., Fanger,G.R., Reltter,M.W., Stolk,J.A., Day,C.H.,  
 Vedrick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.  
 and Henderson,R.A.  
 TITLE Compositions and methods for the therapy and diagnosis of prostate  
 cancer  
 JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;  
 CORIXA CORPORATION (US)  
 FEATURES  
 source 1..444  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 120 a 100 c 78 g 143 t 3 others  
 ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;  
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;  
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 acccagaatccatgctgtaataatggtctcattatccccaagattcttgatgtgcaag 747  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 ACCCAAAATCCAAATGCTGAATATTTGGCTTCATTTCCAGATCTTTGATGTCAAG 60

QY 748 gatttaatgtgtctcagcttgggcactcagttgaagacctaaagatgcccagcag 807  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTTAAGAGATCCAGCGGCAGG 120

QY 808 ttataatgcaagaacaatattcaagcgacacaaggttatggaacttgcccgccagt 867  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 TTTATATATGACAGCAACAATATTCAAGCGGACAGAGTTATTGAATGCGCCAGT 180

QY	407	cactgttagtgatggaagtgaagatttgcacaatccttgaccattgactattatg	466
QY	868	tgaaatttcctccattgacttggaatcctttcaatcagccagagagattgaaatttac	927
Db	181	tgcaatttcattcccatgtgactggagctcttattcattccacccanagagattgaaatttac	240
QY	928	ccctacgactcttctacctcctgagagagccagtggtgtgtagctataagcttgccaca-	986
Db	241	ccctacgactcttcttactctctgagagagccagtggtgtgtagctataagcttgccaca	300
QY	987	tttttttcccttattcctttgtcagagatgtgatcatcatatgactagaacaacag	1046
Db	301	tttttttttcccttattcctttgtcagagatgtgatcatcatatgactagaacaacag	360
QY	1047	agtgactttcacaaaatttcctctagagatgtgataataaaccttacctatagttgcatt	1106
Db	361	agtgacttttacaanaatttctctatagannattgtgataataaaccttacctatagttgcatt	420
QY	1107	acttgctctccctagctatacctc	1130
Db	421	actttgctctccctagctatacctc	444
RESULT	9		
LOCUS	AY029586	1971 bp	mRNA linear ROD 09-MAY-2001
DEFINITION	Mus musculus dudulin 2 mRNA, complete cds.		
ACCESSION	AY029586		
VERSION	AY029586.1 GI:14017412		
KEYWORDS	house mouse, Mus musculus		
SOURCE	ORGANISM Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1971) Serru, V., Lambdin, D., Lenoir, C., Manivet, P., Vaubourdoille, M., Kellermann, O. and Loric, S.		
TITLE	Molecular cloning and expression of mouse dudulin 2		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1971) Loric, S. and Serru, V.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (10-Apr-2001) Biochemistry A Laboratory, Saint-Antoine University Hospital, 184 rue du Faubourg Saint-Antoine, Paris 75012, France		
FEATURES	Location/Qualifiers		
source	1..1971 /organism="Mus musculus" /db_xref="taxon:10090" 46..1590 /codon_start=1 /product="dudulin 2" /protein_id="AAK50539.1" /db_xref="GI:14017413" /translation="MSGEMDKPLISRLVDSGSLAEVPEKAPVGIIGSGDFARSILATRLVQSGVVVGSRNPKRTAGLFPPLSLAQVFOEAVSSEPVIAVAFREHYSLSLSLADOLAGILIVDSNPTEKEHLQHSNAEYIASLPACTVYKANNVISAMALQAVPDLGNROVLIGNSDKAKORYEMARAMEGTPILMGSLASAREVEATDILPLSGKVPDILLALGFLVCEPTYNFIRDVLOPIYTRDENKEFYKMPISVYNTTIPCAVYILSLVYLPQVIAAALQURGRKTYQRPDMLDHLDRKOIGLISFPALHAIYSFCULPARKSHRYDILVNLAVKQYVLANSKRLWEEVEMRMEIYSLGLVLAGMSLAVTSLPSIANSLNKKERSFVAGQVLAFLISTMHILYTGWTRAFEENHYKFLPPTFTLLILPCVILIANKGFLFLPCISRLRTKIRRMENKDAVKFMLPGDTHRGKROATCEALEVEMACGGPELGSLSGCCPAA"		
BASE COUNT	410 a 572 c 548 g 441 t		
ORIGIN			
Query Match	22.0%; Score 415.2; DB 10; Length 1971;		
Best Local Similarity	57.8%; Pred. No. 8.1e-95;		
Matches	760; Conservative 0; Mismatches 548; Indels 6; Gaps 1;		

D	132	CAAAAGTGGGCACTCCTGGGAGGTGGGAGATTTCGCGGCTTCGAGGCAACAAGCGCTGTGGG	191
O	467	atgcgctacacatgctgctacatgagaaagtgaataacccaaagttgcttgaattttcc	526
D	192	CTCTGCTTCAAGTGTGTGTGGGAGCCGTAAACCCCAACACGACGGCTGGCTTCTCC	251
O	527	tcatgigtgtagatgctacatcatataaagtgtctcaaaaaaataataatattgt	586
D	252	CTCCTTACTCTCAATGTACTTTCCAGGAGGAGCCGTGACTCTCCAGAGTCATTTTGT	311
O	587	tgctatacacagagaataattatacctccctgctgtagaacctgtagacatctgctgtagta	646
D	312	GGCCGTTGTCGGGAGAGCACTACTCTCACTGTGCTACTTTGTGTGACCAAGTTGGTCGCA	371
O	647	aatccgattgaatgtagcaataacatgagataaa-----ccagtaaccagaatccaa	700
D	372	GATCCTGTGTGATGTATGAGCAACCCCAACGGAGAGGAGCATTTCTAGCACCCGCACTTAA	431
O	701	tgctgaatattggtctcatattccagatcttctgattgtgccaagaattaaatgtgt	760
D	432	TGCTGAGTACCTGGCTCCTACTCTTTCTCTCGTGCACCTGTGGTGAAGGCGCTTCAAGTAT	491
O	761	ctcagcttgggcaacttaagtaagtaagcaatgccaacggcaggtttatatagtcag	820
D	492	CTCTGCAATGGGCGCTACAGGCGTGTGCCAAGGATGGGAACAGAGTGTCATCTGGG	551
O	821	caacaatatcaagcgcgacaacaagttatlgaaacttgcgcgcsagttgaattcatcc	880
D	552	AAATGACAGCAAGGCCAAMAAGAGTAAATGAGATGAGATGGACCGCCATAGGTTTACACC	611
O	881	cattgacttggagatccctatcatcaacgaagagatlgaaatltacccttagagcttt	940
D	612	CTGTGACATGGGATCCCTGGCTCTACGCGGGAGGTAGAGACCATACCCCTCGCGCTCT	671
O	941	tacctctggagagggccaaatggtagtactaaagcttggccaatttttttccctta	1000
D	672	TCCATCCGGGAAGGTGCCCAACCTCTCGGAGCTGGGGCTTTTGTGTCTTACACTTA	731
O	1001	tctccttgcagagatgtagtcatccatatagtctagaacaacaagatgactttaca	1060
D	732	CACATTCATCCGGGACGCTTCTACAGCCATTAATTCGGAAAGATGAGAACAAAGTTCTACA	791
O	1061	aattccatataagatgtgataataaacttactataagttgcatattacttgcctccct	1120
D	792	GATCCCTTGTCTGTGTGTACACACCACTACCTGTGTGTGTATGTGTCTGTCTCTCT	851
O	1121	agatatactgcagatgtcttcttgcacatgcttaacaattatcttgcacaaagtatag	1180
D	852	GGTGTACTCTCCGCGTGTGCTGTGGCACTGTGGCTTACAGCTCGGAGGGGACCMAATGACA	911
O	1181	gagaattccacactgtgttggaaaacctgtgtacagttgtagaaaacagcttggattactaag	1240
D	912	GCGCTTCCCAAGACTGGCTGAGCAACATGGGTGAGCATGTCAMGACGATCGGGCTGTGAG	971
O	1241	tttttcttcgcatatgtagcaatgtgtcctacagcctcgtgttaaccgatgtagaaggtcaga	1300
D	972	CTTCTTCTTGCGCATGTGTGACGCTCTTACAGCTTCTGTGCGCTGCGCTGCCCTCCCA	1031
O	1301	gagatattgttctcaaaagtgtatacgaagatgtatgaataattgaaaactctg	1360
D	1032	CCGCTACACCTGTGTCATCTGTGATGAGCAGGGTCTCGGCAACAAGACCGGCTCTG	1091
O	1361	gaatgagaaagaagtgttggaaattgaatgtatatactccttggcataatgtagccttgg	1420
D	1092	GGTTTGAGGAAGAACTCTGAGAGATGAGATATACCTGTCTCCGGGTGTGCTGGCCCTGAG	1151
O	1421	cttaacttccctccctggcagtaactctatccctccctgaatgtagcaatgtcttaaacctggag	1480
D	1152	CATGTTGTGCTGCTGTGCTGACCTACCTCGCTCCGTCATCTGCTAATTCCTCAACTGGA	1211
O	1481	agaattcaagttatcatcagctctacaacttggatagatgttgtctgtccataaagtacttcca	1540
D	1212	GGAATTTTAAAGTTGTGTGCAATGCCAATCTGGGCTTGTGTGGCCCTGTGATATCTACCAATAATGCA	1271





|||||  
Db 127745 TCTCAGCTGGGACTTCAGCTAGTCCCAAGATGCCAGCGCCAGGAT 127695  
RESULT 12  
AF335281 2707 bp mRNA Linear ROD 01-PEB-2001  
LOCUS Rattus norvegicus tumor suppressor pHyde mRNA, complete cds.  
DEFINITION AF335281  
ACCESSION AF335281.1 GI:12642828  
KEYWORDS  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 2707)  
Steiner, M.S., Zhang, X., Wang, Y. and Lu, Y.  
Growth inhibition of prostate cancer by an adenovirus expressing a  
novel tumor suppressor gene, pHyde  
Cancer Res. 60 (16), 4419-4425 (2000)  
- JOURNAL MEDLINE  
20424188  
PUBMED 10969787  
REFERENCE  
2 (bases 1 to 2707)  
Lu, Y., Rinaldy, A.R. and Steiner, M.S.  
Direct Submission  
Submitted (08-JAN-2001) Urology, University of Tennessee Health  
Science Center, 956 Court Avenue, H220, Memphis, TN 38163, USA  
FEATURES  
Source  
Location/Qualifiers  
1..2707  
/organism="Rattus norvegicus"  
/strain="Dunning"  
/db\_xref="taxon:10116"  
/cell\_line="At-1"  
/tissue\_type="prostate cancer tumor"  
26..1492  
/codon\_start=1  
/product="tumor suppressor pHyde"  
/protein\_id="AAK00361.1"  
/db\_xref="GI:12642828"  
/translation="MSGDKPLISRLVDSGLAEVPEKAPVGVILGSGDFARSLA  
TRVSGFEVYVNSRNPRTAGLPISLAQVFOEAVSSPEVIVAVFRENYSISL  
ADQLAGILVDVNSPTKEKRLQHSNAEYLAISFPACTVAKANVISAAMLOQPRD  
GNROYLIGDOLLEAKHTVSEMARAMEFTPDGSLASREVEALIRLPLSPMKVPTLL  
ALGISTSVAVNFRDVLQPIYIRDKENKFKYMPISVNTTIPCAVYVLSIVYGLV  
AAALQIRGKRYGKRPMDHMLDHRKQIGLSFFPMALHLYSPCLPDRSHRDLY  
NLAKOVLANKSRLVVEEVRMEITVLSGLVALGMSLAVTSIPSTANSLNKKEFS  
FVOSTTGFVALMISTMTLTLYGWRFAEEENHYKRYLDPPTFLTLPLPCVITLAKGLFL  
LPCLSHRLTKIRGMERDGAVKFMLPAGHTDGEKTSYH"  
BASE COUNT 574 a 788 c 715 g 630 t  
ORIGIN  
Query Match 21.5%; Score 406.6; DB 10; Length 2707;  
Best Local Similarity 57.4%; Pred. No. 1.4e-92;  
Matches 755; Conservative 0; Mismatches 554; Indels 6; Gaps 1;  
Db 407 cactgtatgttgaattggaattttggcaaatcccttgacattgcaattatg 466  
Db 112 CAAAGTGGGATCTCTGGGACGCGGATTTTGGCCGTCCTGGCCACACGCTGTGG 171  
Qy 467 atgcgcatcatgtgcatagaaagtagaaatcctaagttgttctgaattttcc 526  
Db 172 CTCTGGCTTTCTTTGGTGGTGGGAAGCCGTAACCCCAAGACGACGCTGTTCC 231  
Qy 527 tcatgttgtagatgcatcatcatgaagatgcttcacaaanaaataataattgt 586  
Db 232 CTCCTTACCCCAAGTGAATTTCCAGAGAGGCGCTGAGCTCCAGAGGTCTTTGT 291  
Qy 587 tgcatacacagaacattacatccctgtggagacctggaacatctgtgttggttaa 646  
Db 292 GGCGGTGTTCCGGAGGACACTCTCTCACTGTGAGTCTGTGACGAGTTGGTGGA 351

Qy 647 aatccctgattgtgagcaataacatgagataa-----ccagtaccagaatccaa 700  
Db 352 GATTCCTAGTGATGTAGCAACCCACGAGAGAGGCGTCTTACACCGCCAGTCGAA 411  
Qy 701 tgcataatattgcttcattatccagatcttctgattgtlcaaaagattatgt 760  
Db 412 CCGCGAGTACCTGGCTCCCTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471  
Qy 761 ctcaagcttggcaacttcaattgagactaagatgacgacgagcttataatgag 820  
Db 472 CTCTGATGGGCTTACAGGCTGGCCCAAGGATGGGAACAGCAGCTGCTATGCGG 531  
Qy 821 caacaattcaagcgcgcaacaaggtttatgaacttgcgcgcgaattgaatttcc 880  
Db 532 TGACCAAGCTGGAAAGCCACACACACCTCTCAGAGATGGCGGCGCATGGTTAC 591  
Qy 881 catgttgtagatctcatcatcagccagagatgtaaaattacccctagactct 940  
Db 592 ACTGGACATGGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651  
Qy 941 tactctctgagagagccagtggtgtgtagctataaactgtgacacatttttcc 1000  
Db 652 TCCATCTGGAAGGTGCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711  
Qy 1001 ttccttgcagagatgtagtcatcatatgtagtaaaacaaagagtagtcttaca 1060  
Db 712 CAACCTTACATCCGGAGCTTCTACAGCCGTACATCCGGAAGATGACAAAGTTTCA 771  
Qy 1061 aattccatagagatgtgataaanaacttaccatagttgcatcatttgcctcc 1120  
Db 772 GATGCCCTGTCTGTGTGTAACACACAGATACCTGTGTGCTTACGTGCTGTG 831  
Qy 1121 agtatacccgcaagctctctgagctgcttcaacttataagcaccagaatag 1180  
Db 832 GGTATTCGCTGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891  
Qy 1181 ggaattccacacttggttgaaacctggttacaagttagaanaacagcttgat 1240  
Db 892 GCGCTTCCCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951  
Qy 1241 ttttcttgcctatggttcatatggttgcctccttcaagcctcttcaagatga 1300  
Db 952 CTTTCTTTTGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1011  
Qy 1301 gggatattgttcttcaaatggttatacgaaggttcaatgtaaaactctg 1360  
Db 1012 CCGTATGATCTGTGTAACCTGCTGTGAAGCGCTGTGCGCAAGAAGCGCTCT 1071  
Qy 1361 gaatgagagaagatgtgagaattgaaatgtaatactcccttgatgaatag 1420  
Db 1072 GGTGAGGAAGAAGTGTGCGGAGTGAAGATATACATGCTGCTGCTGCTGCT 1131  
Qy 1421 cttacttccctcctgagcaatcttccctcaagtgaagcaatgctttaaactg 1480  
Db 1132 CATGCTGATCAGCTGCGGCGGTATCACTGATCCCTTCAATGCAAACTCACT 1191  
Qy 1481 agaattcagtttattcaatcacttcaacttgatagtcgcctgtgctataag 1540  
Db 1192 GGAAGTACACTTGTGCAATCTGACGCTGGGCTTGGGCTGTGATCTGAGCA 1251  
Qy 1541 tgttaattttagatgagaaagagctttagaagaagtaataagattataac 1600  
Db 1252 CACCTCAGCTACAGCTGAGTACCGGCTTTTGAAGAAACACTACAAGTTCTAC 1311  
Qy 1601 accaacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1660  
Db 1312 ACCCAATTCACAGCTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1371  
Qy 1661 attccttcaatgataagccgaagatgtaaaagaaatgtaaaagcttggaag 1715  
Db 1372 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426

LOCUS	AX211372	3844 bp	DNA	linear	PAT 06-SEP-2001
DEFINITION	Sequence 18 from Patent W00159127.				
ACCESSION	AX211372				
VERSION	AX211372.1	GI:15523741			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 3844)				
AUTHORS	Tang,Y.T., Yue,H., Baughn,M.R., Yao,M.G., Bandman,O., Azimzai,Y., Lal,P., Gandhi,A.R., Ring,H.Z., Shih,L.L., Yang,J. and Pollock,J.L.				
TITLE	Drug metabolizing enzymes				
JOURNAL	Patent: WO 0159127-A 18 16-AUG-2001; Incyte Genomics, Inc. (US)				
FEATURES	Location/Qualifiers				
SOURCE	1..3844				
BASE COUNT	/organism="Homo sapiens" /db_xref="taxon:9606" /note="Incyte ID No: 5629033CB1"				
ORIGIN	862 a 1118 c 1049 g 815 t				
Query Match	20.1%; Score 381; DB 6; Length 3844;				
Best Local Similarity	56.2%; Pred. No. 5.1e-86;				
Matches	739; Conservative 0; Mismatches 570; Indels 6; Gaps 1;				
OY	407 cactgtaggtgtgattgtgaagtgtgaatttgcacaatccttgaccattcgattatag 466				
Db	138 CAAAGTGGGCAATCCCTGGGTAGCGGGAGCTTGGCCCGCTCCGTGGCCACAGCGCTGTGGG 197				
OY	467 atggcgatcatcgtgtgcatagagaagaaatcctaagtttgctctgaatttttc 526				
Db	198 CTCCTGGCTTCAAAAGTGGTGTGGGAGCCGCAACCCCAAGCCACAGCCAGGCTTTTC 257				
OY	527 tcaatgtgttagatgtcactcatcagtaaatgtctctcaaaaaaataataattgt 586				
Db	258 CTCACCGGCCCCAAGAGCATCTTCCAGAGAGAGCATGAGCTCCCGGAGGTGATCTTGT 317				
OY	567 tgcataacacagaagaactatactccctgtgggaacttggagacttgcctgtgtgtgaa 646				
Db	318 GCGTGTGTTCCGGAGCACTACTCTTCCACTGTGCACTGACATGACACAGTGGCGGCA 377				
OY	647 aatcctgattgtatgtgaagcaataaataagataaa-----ccagtaaccagaatccaa 700				
Db	378 GATCCTGGTGTGATGTGAGCAACCCATACAGAGCAAGACACCTTACGATCTGATCCAA 437				
OY	701 tgcgtgaataattgtgctcaattatcccaagattctttagtttcaaaagattatgtgt 760				
Db	438 TGCTAGTACCTGGGCTCTCCCTCTTCCCACTTCACAGTGGTCAAGGGCTTCAAGTCAAT 497				
OY	761 ctacagcttgggaacttgaatagtagaacctaaagatgcagcgcgagttatataag 820				
Db	498 CTCCTCCCTGGAGACCTTCGAGCGCTGGCCCAAGGATGTAAACAGCAGGTGCCATGTGGG 557				
OY	821 caacaataattcaagcgacaacagattatgaacttgcagctgcgcagttgaattatcc 880				
Db	558 TGACCAAGCCAGACCAAGCGTGTCTCGGAGATGAGCGCTCGCATAGGGCTTCATGCC 617				
OY	881 catgaacttggatctctatatactacgacagagatttgaataattaccctcagaactct 940				
Db	618 CGTGAACATGGGATCCCTGGCTCAAGCTCGGAGGTGAGGCGCAAGCCCTGCGGCTCTCT 677				
OY	941 tactctcgtgaagagcgacagtgtgtgtatgataagcttgcagcaacttttttccctta 1000				
Db	678 CCGCGGCTGGAGAAAGTGGCCACCGCTCTGGGCGCTTTCGTCGTCTGTATAGCTTA 737				
OY	1001 ttctcttgcagagatgtgatcatcaatagtctagaacaacaacagagtgcatttaca 1060				
Db	738 CAATCTCTCCCGGGACGCTTTCGACCCCTATGTGAGGAAAGGCCGAACAAGTTCCTCA 797				

Oy	1061	aattccctataagaagcttgatggaataaaaccttaccctatagttggtccattacttgccttcct	1120
Db	798	gctgcgccggtgcgcgggtccaaacacccacacacacccgctgcggcctgacgtgctgcttccact	857
Oy	1121	agatacctcgcgaagctctctctgcacgcctgcttatacacttatacaggcaccacaagtatag	1180
Db	858	cgtgtaccttggcccccggcgcgtgcgtggcgctgcccctgcagctggccggccgaccacaaattacca	917
Oy	1181	gagatttccaccttggttggaaaacctggtttacagttgttgaanaacagcttggattactaag	1240
Db	918	gcgcgttccccccgactgcgtgcgacacacatggcctacagcaccgcgaagcagatccggcctgctcag	977
Oy	1241	ttttctctgctatagfcccactgltgtccctaaagcctctgcgtcttaccgagatgagaagttaga	1300
Db	978	cttcttcttgcgcgcgcgccttcacacgcccctctacagcttctgcttgcctgcgcgcgcgcgccca	1037
Oy	1301	gagataattgtttcttccacaatagctgttaccagcaggtttcaatgcacaatatgaaacactctg	1360
Db	1038	ccgctacgacacccgtgcac	1097
Oy	1361	gaatgaggaagaagaagttgagagaattgaaatgatactctcctcttggcataatgagccttgg	1420
Db	1098	ggatgagagagagagggctctgcgcgatggaatctacaccttccctggagagctgcgcgcctcgg	1157
Oy	1421	cttaacttccccctcgcgcgaatcacttcatcccttcagtgagcaatgctttaaacttggag	1480
Db	1158	cacgtgttgcctcctcgtgc	1217
Oy	1481	agaattcagatttatacagcttcaacacttgatgataitgctcgtcgtcctacataagtaacttcca	1540
Db	1218	ggagtttcagcttgcgttcagctcctcactgsggcttggcgccctcgctgagacacactgcca	1277
Oy	1541	tgtttaattatgatgatggaacagagccttggagagaagtaactacagatttataacc	1600
Db	1278	cacgcttccacacacgcttgc	1337
Oy	1601	accaaactctgtctctgtcgtctgttgccttcccaatgatacttctggtgataattt	1660
Db	1338	tcgccaccttccacgcgtcacacgcgtgctgctgctcctcctgcgtctacatcctggcccaaacctctgt	1397
Oy	1661	attcctccatgataaagccgaagaacttcaaaaacgaattaaaaagccttggagaag	1715
Db	1398	tctcctgcctgcacatcacccgac	1452
RESULT 14			
LOCUS	AY029585	1845 bp	mRNA
DEFINITION	Homo sapiens dudulin 2 mRNA, complete cds.		
ACCESSION	AY029585		
VERSION	AY029585.1	GI:14017410	
KEYWORDS	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Serru,V., Manivet,P., Lenoir,C., Eschwege,P., Lambin,D., Vaudour-Dolle,M., Kellermann,O. and Loric,S.		
TITLE	Dudulin 2, a new tumor antigen expressed in various human tumors		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1845)		
TITLE	Loric,S. and Serru,V.		
JOURNAL	Direct Submission		
SOURCE	Submitted (10-APR-2001) Biochemistry A Laboratory, Saint-Antoine University Hospital, 184 rue du Faubourg Saint-Antoine, Paris 75012, France		
FEATURES	Location/Qualifiers		
SOURCE	1..1845		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="2"		

CDS	135..1601
	/codon_start=1
	/product="dudulin 2"
	/protein_id="AAK50538.1"
	/db_xref="GI:14017411"
	/translation="MPEMDKPLISLHVDSDSLAVPDEAPKVGILGSGDFARSILA TRIGSGPKVYVSSRNPKRTARLYPSAAQYTFQDEAVSPPIVFNAPREYSSLSLSI SDLAGKILVDVSNPTFQEHLOHESNAEYIASLPPICTYKATNVIISAWLQAPKPK GNRPVPLCGDQPEAKRAVSEMLAMGMPMDVMSVNTLPCVAYLISLVPYPLIL ALGVEFYADNPFRADVLOPYVDSQSKFELPVSIVNTLPCVAYLISLVPYPLIL AAALQRRGTRKYQRPEDMLDHLQHRKQIGLSPFCALIAHLVIFQCLPRAHRYDLY NLAKVOLANKSHLMVDEEVRMEIYVSLGVALAGTSLAVTSPLSIANSIMHREPS FVQSLSEFVALVLTSLHTITRYGMRAPDEESYKRYLPLPTITLIVCVILANALFEL LPICSRRLARIRRMERBSSTIKTFLPDLHALERTSHV"
BASE COUNT	349 a 591 c 523 g 382 t
ORIGIN	
Query Match	20.1%; Score 379.4; DB: 9; Length 1845;
Best Local Similarity	56.1%; Pred. No. 1,1e-85;
. Matches 738; Conservative	0; Mismatches 571; Indels 6; Gaps 1;
Db	407 cacttagtgatgattggaatggaattttgccaatcccttgaccattgactatag 466
Db	221 CAAAGTGGGCACTCCGGGTAGCGGGGACTTTGCCCGCTCCGACACAGCCTGTGG 280
Db	467 atgcgcatcaatggtgcatagagaagtagaataatcctaagttgtcttgaattttcc 526
Db	281 CTCTGGCTTCAAGTGGTGGGGAGCGGACGCCAAGCCAAAGCAGCAGCGCTGTATCC 340
Db	527 tcatgtgttagatgtcactcatcatgaaagtcttcaaaaacaataataatttgt 586
Db	341 CTCACGGCCCAAGAGCACTTTCCAAAGAGAGGCACTGAGCTCCCGGAGGTCATCTTGT 400
Db	587 tgcctaacagagaacatttacctccctcttgggaacttgagaactctgctgtgtgttaa 646
Db	401 GGCTGTGTTCCGGGAGCACTACTCTTCACTGTGACGTCTCACTGACCACTGGCGGCA 460
Db	647 aatcctgattgatgtgagcaataaactagagtaaa-----cagtaaccagaattcaa 700
Db	461 GATCCTGGTGTGATGAGACCAACCTTACAGAGCAAGAGCAACCTTTCAGCATTCGTGATCCAA 520
Db	701 tgcctgaataattgctctcaattatcccaagattctttgattgtcaagaagattatgtgt 760
Db	521 TGTGTAGTACCTGGGCTCTCTCTTCCCTCCACTTGTGACACATGTGTCAGGCTTCAATGTAT 580
Db	761 ctgaagctggagcacttgaagttagagacccaagaagtgcagccggcaggttatatatgca 820
Db	581 CTCTCCCTGGGACCTGTGACGGCTGGGCCCAAGGAGTGTAAAGGCGGTGCCATCTGTGGG 640
Db	821 caacatatatcaagtcgagacaacaagttatgnaacttgcgcgcagcttgaatttcatcc 880
Db	641 TGACCAAGCCAGAGCAAGCCGCTGCTCGAGATAGGCGCTCCGCATGGGGCTTCATGCC 700
Db	881 catgtaacttggatcctatcatcagccagagagaattgaaatttacccttaccagactct 940
Db	701 CGTGCACATGGAATCCTGCGGCTGACGCTGGAGAGTGAAGGCGCATGCCCTGCGGCTCCT 760
Db	941 tactctctggagagggccagatggtgtgtgtgataagcttggccacatatlttttccctta 1000
Db	761 CCGGCGCTGGGAAGGCCACCCCTGTCTGGCCCTGGGCGCTTTCGTCCTCTTATGCTTA 820
Db	1001 ttctcttgcagagatggtgattcatcatcatatgctagagaaccaacagagtgacttttcaa 1060
Db	821 CAACCTTCTCCGGGACGCTTGTGACGCCCTATGTGACGAAAGCCAGCAACAGTTTTCAA 880
Db	1061 aattcctatagagatttgaataaaaccttacctatagttgccaattacttcttccct 1120
Db	881 GCTGCGCTGTGCCGGTCAACACACACTGCGCGGTGCTACGTGCTGTACT 940
Db	1121 agtatactcgcaagctctcttggcagcgtcttatcaacttattatcagcaccagaattag 1180
Db	941 CGTGTACTTGGCCGGCGGTGGCGGCTCTGTGAGCTGGCGCGGCCACCAATGATCCA 1000

FEATURES	Source
ORGANISM	Location/Qualifiers
REFERENCE	1. 3912
AUTHORS	Isogai,T. and Otsuki,T.
TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT	NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5' - 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites)
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
JOURNAL	NEO human cDNA sequencing project
COMMENT	unpublished (2000)
LOCUS	AK001691 3912 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10829 fis, clone NT2RP400118.
ACCESSION	AK001691
VERSION	AK001691.1 GI:7023106
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP400118.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; E



```

/db_xref="taxon:9606"
/clone="NT2RP4001138"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_1ib="NT2RP4"
/notes="cloning vector: pME18SFL3: mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
135..1601
CDS
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91839.1"
/db_xref="GI:7023107"
/translation="MPEEMDKPLISLHLVSDSLSAKVPDEAPKVGILSGDFARSLSA
TRLVSGFKVVGSRNPKRTARLPSAQTVOEAVSSPEVTFVAVFRRHYSLSLCSL
SDLAGKILVDNPTDEHLOHRESNAIYLAIPCTCVVACNFIYASMTLOAGPRD
GNROYDIDGDEAPKRAVSEMAIANGPMVDGSLASAEVEMPRLLIPANKVPITL
ALGLFVCPYAVNFPVNDVLOPYQESONKFEKFPYSVNTTLCVAVYLLSVLYRVL
AAALDRGKTQRFPPDWLHMLQHKQJGLISFCAALHLYSFCULPRKARIDLV
NLAVQVLANKSHLWVEEVEWMEIYLSGLVALGTLISLAVTSLSINSINWREFS
FVOSLSGFVALVSLHLTLTGMTAFESRYKFLPPTFTLLVPCVYIILAKLFL
LPCISRRILARIRGWERESTIKFTPTDHALAEKTSHV"
BASE COUNT      863 a      1148 c      1071 g      830 t
ORIGIN
Query Match      20.1%; Score 379.4; DB 9; Length 3912;
Best Local Similarity 56.1%; Pred. No. 1.3e-85;
Matches 738; Conservative 0; Mismatches 571; Indels 6; Gaps 1;
QY 407 cactgagtgatgattggaagtggaatttgcacaatccttaccatctgactttag 466
    || || || || || || || || || || || || || || || || || || || ||
Db 221 CAAAGTGGGCACTCCGTAAGCGGAGCACTTGGCCCTCCCTGCGCACACAGCTGGGG 280
    467 atcgagctacatgltgtaagaagtagaaatccctaaattgcttctgaattttc 526
    || || || || || || || || || || || || || || || || || || || ||
Db 281 CTTGTGCTTCAAGTGTGGTGGGAGCCGCAACCCCAAGCAGACGCGCTGATCC 340
    QY 527 tcatgtgtagatgtaactcatatgaagatgcttcaaaaaaataataatttgt 586
    || || || || || || || || || || || || || || || || || || || ||
Db 341 CTCAGCGGCCCAAGTCACTTCCAAAGAGGAGGAGTCCCGGAGGTCACTTTTG 400
    QY 587 tgtataacagaagaatataatccctctgtgagacctgaagacatctgttgtgtaa 646
    || || || || || || || || || || || || || || || || || || || ||
Db 401 GGTGTGTTCCGGGACACTACTTCTACGTCTCAGTGCAGTCCGAGGCGGCA 460
    QY 647 aatcctgattgattgagaataacatgagaataa-----ccagtaccagaatccaa 700
    || || || || || || || || || || || || || || || || || || || ||
Db 461 GATCCGTGGTGGATGTAGACCACTTACAGAGCAAGACACTTACAGCATGTAGTCCA 520
    QY 701 tgcgaataattggtctcatattccacagattcttgattgtcaagaattatttgt 760
    || || || || || || || || || || || || || || || || || || || ||
Db 521 TGCTGAGTACCTGCTCCCTCTTCCCACTTGCACAGTGTCAAGGCTTCAATGTAT 580
    QY 761 ctacgttggaccttcaagttgaagtaagatgcacgagcagttatataatgcaag 820
    || || || || || || || || || || || || || || || || || || || ||
Db 581 CTCTGCTGACCTCAGAGCTGAGGAGGATGTAAACAGGAGGTGCCATCTGCGG 640
    QY 821 caacaatttcaagcgcgaacaggttattgaactgcccgcagttgaattatcc 880
    || || || || || || || || || || || || || || || || || || || ||
Db 641 TGACCAAGCCAGAGCCAGAGCTGCTGTCTGAGATGGCGCTGCCATGGGCTTCAATGCC 700
    QY 881 catgacttggagatccttcatcagcagagagattgaataatttaccctacgactct 940
    || || || || || || || || || || || || || || || || || || || ||
Db 701 CGTGACATGGGATCCCTGGCGCTGACCTGGAGGATGGAGGCCATGCCCTGCCCTCT 760
    QY 941 tactctctgagagagccagtggtgtagctataaagcttggccaatlttttctctta 1000
    || || || || || || || || || || || || || || || || || || || ||
Db 761 CCCGGCTGGAAGTGCACACCTGCTGGCCCTGTTGTGCTTCATGCTTATGCTTA 820
    QY 1001 ttccttgcagagatgattcatcatatgtagaaacaaagagatgactttaa 1060
    || || || || || || || || || || || || || || || || || || || ||
Db 821 CAACCTTCGTCGGGACGTTCTGACGCCCTATGTGACAGAAAGCCAGAACAGTTTCA 880

```

```

QY 1061 aattcctatagagattgtgataaataaacttaccatagattgcaattacttctccct 1120
    || || || || || || || || || || || || || || || || || || || ||
Db 881 GCTGCCCGGTGCTCGGTGTCAACACACACATGCGGTGGGCTTACGTGCTGCTACT 940
    QY 1121 agatatactgcagagcttcttgcagagctgttatacattatataagcagcaagtag 1180
    || || || || || || || || || || || || || || || || || || || ||
Db 941 CGGTACTTCCCGCGCTGCTGGCGCTGCTGACCTGCGGCGGCGGCACTGACCA 1000
    QY 1181 gagatttcaacctgtgtggaacactgttaccagtgtagaanaacagcttgattactaag 1240
    || || || || || || || || || || || || || || || || || || || ||
Db 1001 GCGCTTCCCGGACTGCTGAGAACCACTGCTACACACACAGCAAGATGCGGCTGAC 1060
    QY 1241 tttttctgcatagtgttgcacagcttgcacagcttgcatacgaatgagaagtcaga 1300
    || || || || || || || || || || || || || || || || || || || ||
Db 1061 CTCTCTGCGCGCGCCCTGCACGCGCTCTACAGCTTCTGCTGCGCTGCGCGCGCCA 1120
    QY 1301 gatatatttcttcaacttgcatacgaagcttcaagcaatcttgaanaactcttg 1360
    || || || || || || || || || || || || || || || || || || || ||
Db 1121 CCGCTACGACCTGCTCAACCTGCGCAGTCAAGCAAGGTCTTGCGCAAGCAAGCA 1180
    QY 1361 gaatgaagagaagatttgaagaattgaatgatatccttcttgcataatgagccttgg 1420
    || || || || || || || || || || || || || || || || || || || ||
Db 1181 GGTGAGAGGAGAGGTCTGCGGAGATGAGATCTACCTCTCCCTGGAGATGCTGGCCCTCG 1240
    QY 1421 cttaattccctctgcagcagcttctatcccttcaagtagaagaatgtttaaactggag 1480
    || || || || || || || || || || || || || || || || || || || ||
Db 1241 CAGGTTTCCCTGCTGCGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1300
    QY 1481 agaatcagtttattatagcttcaacttgaatgcttgccttgccttgaatgacttcc 1540
    || || || || || || || || || || || || || || || || || || || ||
Db 1301 GGAATTGAGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1360
    QY 1541 tgtttaattatgatatgagaacagagcttlttgaaggaagtagtacaagatttatacacc 1600
    || || || || || || || || || || || || || || || || || || || ||
Db 1361 CAGGCTACCTAGCGGTGAGACCGCGCTTGAGAGAGCGGATCAAGTTTACCTAGTCC 1420
    QY 1601 accaacttgttcttgccttgttcttgccttcaatgtatcttggtagaattattt 1660
    || || || || || || || || || || || || || || || || || || || ||
Db 1421 TCCACACTTACGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
    QY 1661 atctctcattatgataagcgaagaagcgaatgaatgaatgaatgaatgaatgaatga 1715
    || || || || || || || || || || || || || || || || || || || ||
Db 1481 TCTCTGCTCCTGCATCAGCCGACACTGCGCAGAGATCCGAGAGGCTGGAGAGG 1535

```

Search completed: May 9, 2002, 02:12:49  
 Job time: 18694 sec



; PAGE# 0529303

```

: GENERAL INFORMATION:
: APPLICANT: Daniel E. Afar
: APPLICANT: Rene S. Hubert
: APPLICANT: Kahan Leong
: APPLICANT: Arthur B. Raitano
: APPLICANT: Douglas C. Saffran
: APPLICANT: Steve Chappell Mitchell
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: FILE REFERENCE: 129.16US02
: CURRENT APPLICATION NUMBER: US/09/323,873A
: CURRENT FILING DATE: 1999-06-01
: PRIOR APPLICATION NUMBER: 60/087,520
: PRIOR FILING DATE: 1998-06-01
: PRIOR APPLICATION NUMBER: 60/091,183
: PRIOR FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 339
: TYPE: PRT
: ORGANISM: Homo Sapiens
US-09-323-873A-2

```

```

? APPLICATION NUMBER: US/09/083,521
? FILING DATE: Herewith
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: CERRONE, MICHAEL C.
? REGISTRATION NUMBER: 39,132
? REFERENCE/DOCKET NUMBER: PF-0527 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (650) 855-0555
? TELEFAX: (650) 845-4166
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 141 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: PROSTUT10
? CLONE: 1691243
? US-09-083-521-1

```



```

/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: William T. Han
/ REGISTRATION NUMBER: 34,344
/ REFERENCE/DOCKET NUMBER: A7G50022
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 610 270 5024
/ TELEFAX: 610 270 5090
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 365
/ TYPE: Amino Acid
/ TOPOLOGY: Linear
/ US-08-724-974A-3

Query Match          4.4%; Score 112.5; DB 2; Length 365;
Best Local Similarity 20.6%; Pred. No. 0.0025;
Matches 56; Conservative 46; Mismatches 93; Indels 77; Gaps 11.

QY 255 VKTTLPIATITLLSVLYAGLLAAAYOYGYGTKKRRFPMTLWLOCKOKGLISFFPA 313
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 19 HQTLPARYVYYVL-----VGFRPANCLSIYFG-----YLQ--IKANELGVILCNT 64
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 314 MYAVAYSICLPKRRSERYFLNMAYQQVHANIENTSNWEERWRIEMYISFGIMSLGLSL 373
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 65 VADLFYICLSLPFW-----LQVLDHDMNSHGDLSCVCGILLYENIYL-SVGFLCC 114
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 374 LAVTSIPSSNMLNMRFEFSIOSTLGVA-----LIISTF----- 408
    ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 115 ISVDRLAYAHPPRPHQFTLKAAARVTVIWAKKELLTIYFLMHEEVIDENQHRCPE 174
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 409 HVLIYGMRKAFEEHYRYFTPNPFVALVLPST-----VILG 445
    |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 175 HPIIDAMQAI--NYIRFLVGFLFPCILLASTYOGILRAVRSHGTQSRKDQIORLVLS 232
    ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 446 KIIFLPCL-SRKLRINKGEKSQPLEGI 475
    ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 233 TVVIFLACFLPYHVLLVLRVSWEASGDFAKGV 264
    ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT      8
US-09-323-873A-20
/ Sequence 20, Application US/09323873A
/ Patent No. 6329503
/ GENERAL INFORMATION:
/ APPLICANT: Daniel E. Afar
/ APPLICANT: Rene S. Hubert
/ APPLICANT: Kahan Leong
/ APPLICANT: Arthur B. Raitano
/ APPLICANT: Douglas C. Saffran
/ APPLICANT: Steve Chappell Mitchell
/ TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
/ FILE REFERENCE: 129.16USU2
/ CURRENT APPLICATION NUMBER: US/09/323,873A
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 20
/ LENGTH: 34
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: STEAP-1 PEPTIDE
US-09-323-873A-20
```

```

Query Match Similarity      4.2%; Score 107; DB 4; Length 34;
Best Local Similarity      58.8%; Pred. No. 0.00025;
Matches      20; Conservative      5; Mismatches      9; Indels      0; Gaps      0;

QY      326 RRSRYRLFLNMAVYOQVHANIEINSNDEEVWRIEM 359
        ||| || | |||| | : : | : ||| : |
Db       1 RRSRYRKLNLMAVYOQVKNEKEDAWIEHWRMFI 34

RESULT      9
US-08-487-886-2
; Sequence 2, Application US/08487886
; Patent No. 5744448
GENERAL INFORMATION:
APPLICANT: Kelton, Christie Ann
APPLICANT: Schweickhardt, Rene Lynn
APPLICANT: Cheng, Shirley Yui Yen
APPLICANT: Nugent, No. 5744448een Patrice
TITLE OF INVENTION: Human Follicle Stimulating
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
ADDRESSEE: Ares-Serono, Inc.
STREET: Exchange Place, 37th floor
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
COMPUTER: IBM PS/2, model 55 SX
OPERATING SYSTEM: MS-DOS version 4.0
SOFTWARE: VAX/VMS Massll via Kermit to IBM MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/670,085
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Stephan P.
REGISTRATION NUMBER: 28546
REFERENCE/DOCKET NUMBER: US/252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 723-1300
TELEFAX: (617) 723-8923
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 695
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: signal sequence
LOCATION: -17 to -1
IDENTIFICATION METHOD: hydrophobic
FEATURE:
NAME/KEY: putative amino-terminal extracellular domain
LOCATION: 1 to 349
IDENTIFICATION METHOD: similarity with other
IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
IDENTIFICATION METHOD: domains, hydrophilic
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 350 to 613
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
FEATURE:
NAME/KEY: putative transmembrane region I

```

LOCATION: 350 to 370 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region II  
LOCATION: 382 to 404 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region III  
LOCATION: 427 to 448 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region IV  
LOCATION: 469 to 491 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region V  
LOCATION: 512 to 533 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region VI  
LOCATION: 557 to 580 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative transmembrane region VII  
LOCATION: 592 to 613 similarity to other G  
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
FEATURE: putative carboxy-terminal intracellular  
NAME/KEY: domain  
LOCATION: 614 to 678  
US-08-487-886-2

Query Match 4.2%; Score 106; DB 1; Length 695;  
Best Local Similarity 16.4%; Pred. No. 0.031;  
Matches 84; Conservative 81; Mismatches 148; Indels 198; Gaps 18;

QY 79 HEDALTKITIVVAIHREYTSMLDLRLVVG-----KIILDSNNMR 121  
DB 98 HEIRIEKANNL-LYINPEFOMLPNLOYLLISNTGIKHLPDVHKIHSLOKVLIDIDNIN 156  
QY 122 INQVESNAEYLASFPPDLIVKGFNVVSAMALQGPDKASROYICSNINIAQOVIEL 181  
DB 157 IH-----TIERNFVGLSFESEYILML-----NKGIOEINCA----- 189  
QY 182 AROLNFIPIIDGLSSAREIENLPLRLFTLMRGPPVVAISLATFFELVSFVDVIHPYAR 241  
DB 190 ---FNGTQDELINLSDNNLLELPNDVFGASGPVILDISPTRISLPSYGLLENKILUA 246  
QY 242 NOOSDEYKIPRIIVKTL-PIVAITLLSLVY----- 271  
DB 247 RSTYMLKLP-----TEKIVLMEASLTYPSCAFANMRQISELPICKNSILROE 300  
QY 272 -----LAGLIAAAVOLYGYKRRF-----PPMLFTWLOCRQL 305  
DB 301 VQYMTQTRGRRSLAEDNNSYSRGEDMTYEFDDLCNEVVDYICSPFPAFNCEDIM 360  
QY 306 G-----LISFFAMVVA-----YSLCLPMRSERYLFLNMAVQOVHANEN 347

DB 361 GYNILRLIMFISILATIGNIIVLITTSQYKLTVP-----RFLMCNLAFAAD----- 408  
QY 348 SNNEEVWNIEMVISMGSIMSLGLSLAATSIPSVNAINMRE-----SFISQSTL 398  
DB 409 -----LCIGIYLLITASVDIHKTSOYHNVAIDMORGACDAGFPTVASEL 455  
QY 399 G---YVALLISFPH-----VLIYGMKRAFEERYRPTPNFV-LA 435  
DB 456 SVYTLTAITLERHNTTHAMQDCKVOLRHAASVWNGMIFAFALFLPFGISSYMKVS 515  
QY 436 LVLP-----SVIILGKIILPLPC 453  
DB 516 ICLPMDISPLSOLYWSILVILNVLAFVVIC 546  
RESULT 10  
US-08-482-855-2  
; Sequence 2, Application US/08482855  
; Patent No. 6121016  
; GENERAL INFORMATION:  
; APPLICANT: Kelton, Christie Ann  
; APPLICANT: Schweickhardt, Rene Lynn  
; APPLICANT: Cheng, Shirley Vui Yen  
; APPLICANT: Nugent, No. 6121016een Patrice  
; TITLE OF INVENTION: Human Follicle Stimulating  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Stephan P. Williams,  
; STREET: Ares-Serono, Inc.  
; CITY: Exchange Place, 37th floor  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density  
; COMPUTER: IBM PS/2, model 55 SX  
; OPERATING SYSTEM: MS-DOS version 4.0  
; SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,855  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/670,085  
; FILING DATE: 15-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Stephan P.  
; REGISTRATION NUMBER: 28546  
; REFERENCE/DOCKET NUMBER: US/252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 723-1300  
; TELEFAX: (617) 723-8923  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 695  
; TYPE: Amino acid  
; TOPOLOGY: Linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: signal sequence  
; LOCATION: -17 to -1  
; IDENTIFICATION METHOD: hydrophobic  
; FEATURE:  
; NAME/KEY: putative amino-terminal extracellular domain  
; LOCATION: 1 to 349  
; IDENTIFICATION METHOD: similarity with other  
; IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular  
; FEATURE:  
; NAME/KEY: transmembrane domain  
; LOCATION: 350 to 613

```

1 IDENTIFICATION METHOD: similarity to other G
2 IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
3 FEATURE:
4 NAME/KEY: putative transmembrane region I
5 LOCATION: 350 to 370
6 IDENTIFICATION METHOD: similarity to other G
7 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
8 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
9 FEATURE:
10 NAME/KEY: putative transmembrane region II
11 LOCATION: 382 to 404
12 IDENTIFICATION METHOD: similarity to other G
13 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
14 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
15 FEATURE:
16 NAME/KEY: putative transmembrane region III
17 LOCATION: 427 to 448
18 IDENTIFICATION METHOD: similarity to other G
19 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
20 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
21 FEATURE:
22 NAME/KEY: putative transmembrane region IV
23 LOCATION: 469 to 491
24 IDENTIFICATION METHOD: similarity to other G
25 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
26 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
27 FEATURE:
28 NAME/KEY: putative transmembrane region V
29 LOCATION: 512 to 533
30 IDENTIFICATION METHOD: similarity to other G
31 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
32 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
33 FEATURE:
34 NAME/KEY: putative transmembrane region VI
35 LOCATION: 557 to 580
36 IDENTIFICATION METHOD: similarity to other G
37 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
38 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
39 FEATURE:
40 NAME/KEY: putative transmembrane region VII
41 LOCATION: 592 to 613
42 IDENTIFICATION METHOD: similarity to other G
43 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
44 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
45 FEATURE:
46 NAME/KEY: putative carboxy-terminal intracellular
47 NAME/KEY: domain
48 LOCATION: 614 to 678
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
92
```

[illegible]

```

Db      301  VDMYQTFGGQSSSLAEDNESSYSGFDMYTFTEPFYDLCNEVVDTCSPKPDAAFPCEIDM 360
QY      306  G-----LLSFFFAVHVA-----YSDCLPMSRSERYFLNMAVQGVHANIEN 347
Db      361  GYNILIRLWIFSLTALGTNIIVLVITTSQYKLTVP-----RFLMCKLAFAD----- 408
QY      348  SWNEEEVWRIMYISFGIMSLGLSLTAVTSIPSVSNALMRE-----ESFIOSTL 398
Db      409  -----LCIGIYELLIASVDIHRKSOYHNAYIMQAGCADAAGPFTVASEL 455
QY      399  G---YVALLISTFH-----VLIYGMKRAAEERYYRTTPPNFY-LA 435
Db      456  SVYTLTATLTERMHTITHAMQDCKVQLRHAASVYMGWGLFAAPAALEPIFGISSYKWS 515
QY      436  LVLP-----SIVLGRKIIILFLPC 453
Db      516  ICLPMDIDSPLSQLYKMSLVLANTLAFVVIC 546

```

RESULT 11  
 US-09-316-083-3  
 : Sequence 3, Application US/09316083A  
 : Patent No. 6280942  
 : GENERAL INFORMATION:  
 : APPLICANT: The Institute of Physical and Chemical Research  
 : TITLE OF INVENTION: Endonuclease  
 : FILE REFERENCE: PH-651  
 : CURRENT APPLICATION NUMBER: US/09/316,083A  
 : CURRENT FILING DATE: 1999-05-20  
 : EARLIER APPLICATION NUMBER: JP98/141861  
 : EARLIER FILING DATE: 1998-05-22  
 : NUMBER OF SEQ ID NOS: 38  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 3  
 : LENGTH: 476  
 : TYPE: PR1  
 : ORGANISM: Saccharomyces cerevisiae  
 : US-09-316-083-3

Query Match	4.0%;	Score 101;	DB 4;	Length 476;
Best Local Similarity	20.7%;	Pred. No. 0.057;		
Matches 52;	Conservative 48;	Mismatches 85;	Indels 66;	Gaps 12

  

QY	62	NPKEASEEPFPHVVDYTHHEDALFTKTNI----	LEVAIHREHYT-----	SLMD	103
		: : : : :     : : : : :	: : : : :		
Db	254	NPYEVNATSIIN-----	KTNLAKKEKIFETIYNKLYSIDYKIQINNHIPYNYLK		302
QY	104	LRHLVGLIILIDVSNMNRINQY	PESNAEYIASLF--	PSLIYKGRNVVSAMALQIGPKDA	161
		: : : : : : : : : : : : : : : : :	: : : : : : : : : : :		
Db	303	INNKLPINIMDIKKNWYLAGFTAADGSLSSWMPKDTLTKNM-----			347
QY	162	SROYVICSNNIOAROOVETLAROLNF	IFIDIGLSLSAREIENLPRLFTLMRGPVVAIS		221
		:   :   :   :   :   :   :   :   :	:   :   :   :   :   :   :		
Db	348	RPSYVIS-QVETRKELIYLQ-----	SPDL-SISNVKKVGRKLKDFLFLRTIDELMK		400
QY	222	LATEPELISFVRDVIHPARNOOSFYRIP----	LEIYNKTIPLIYAITLISLYIAGLIA		277
		:   :   :   :   :   :   :   :   :	:   :   :   :   :   :   :		
Db	401	-----FIYFF--DKELPLDNKQFNYIKERENFTIKSYMMNRVGLVLS--	YINNKI		451
QY	278	AAQOLYVGTRY			288
Db	452	DNYDIYYINKY			462

  

RESULT	12
US-08-460-576-2	
Sequence 2, Application us/08460576	
Patent No. 6033903	
GENERAL INFORMATION:	
APPLICANT: SISK, William P.	
TITLE OF INVENTION: A METHOD OF EXPRESSING AND SECRETING	
TITLE OF INVENTION: SOLUBLE EXTRACELLULAR DOMAINS OF HUMAN GONADOTROPIN	

Thu May 9 08:44:29 2002

us-09-802-520-1.raii

Page 7

```

1  TITLE OF INVENTION:  HORMONE RECEPTORS
2  NUMBER OF SEQUENCES:  18
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  BROWDY AND NEWMARK
5  STREET:  419 Seventh Street, N.W., Suite 300
6  CITY:  Washington
7  STATE:  D.C.
8  COUNTRY:  USA
9  ZIP:  20004
10
11  COMPUTER READABLE FORM:
12  MEDIUM TYPE:  Floppy disk
13  COMPUTER:  IBM PC compatible
14  OPERATING SYSTEM:  PC-DOS/MS-DOS
15  SOFTWARE:  Patent In Release #1.0, Version #1.30
16  CURRENT APPLICATION DATA:
17  APPLICATION NUMBER:  US/08/460,576
18  FILING DATE:
19  CLASSIFICATION:  435
20  ATTORNEY/AGENT INFORMATION:
21  NAME:  YUN, Allen C.
22  REGISTRATION NUMBER:  37,971
23  REFERENCE/DOCKET NUMBER:  SISK-1
24  TELECOMMUNICATION INFORMATION:
25  TELEPHONE:  202-628-5197
26  TELEFAX:  202-737-3528
27  TELEX:  248633
28  INFORMATION FOR SEQ ID NO:  2:
29  SEQUENCE CHARACTERISTICS:
30  LENGTH:  390 amino acids
31  TYPE:  amino acid
32  TOPOLOGY:  linear
33  MOLECULE TYPE:  peptide
34  OS-08-460-576-2

```

```
Query Match          3.8%; Score 97.5; DB 3; Length 390;  
Best Local Similarity 21.3%; Pred. No. 0.097;  
Matches    45; Conservative   39; Mismatches   72; Indels    55; Gaps      7.  
  
Qy       79 HEDALTKTNIFVALIHREHYTSLMDLRHLVY-----KILIDVSNNMR 121  
           || : | : | : | : | : | : | : | : | : |  
Db        122 HEIRIEKANNT-LYNINPAFOMLPRIQLHLSINTGIKHLPDVHKIHSLOKLVDIGDINN 180  
  
Qy       122 INQPESNAEVLASLFPDSLYVGKGNYVASAMALOGLPDKASQOVIYC>NNICAROOYTEL 181  
           | : :: : | : | : | : | : | : | : | : | : |  
Db        181 IH-----TIERSSEVGLSPESVTLWG---NKNNGIOELHNCA----- 213  
  
Qy       182 AROLNFIPIDGSSLAAREIENLPRLPTLMGPVVAVIASLTFFEFISFDVIIHPYAR 241  
           | - | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db        214 ---FNQGLODELNLSDNNNELPDVHGASGVPILDISRHRHSLPSYGLENKKLLRA 270  
  
Qy       242 NQOSDFYKIPIETIVKYTL-PIVAITLSLVY 271  
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db        271 RSTYMLKKLP-----TLEKLVAMEASTLY 295  
  
RESULT     13  
US-07-757-342D-6  
Sequence 6, Application US/07757342D  
Patent No. 6218509  
GENERAL INFORMATION:  
APPLICANT : IGARASHI, Masao  
            MINEGISHI, Takashi  
            NAKAMURA, Kazuo  
TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
             CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA
```

```

1      ZIP: 02109
2
3      COMPUTER READABLE FORM:
4
5      MEDIUM TYPE: Floppy disk
6
7      COMPUTER: IBM PC compatible
8      OPERATING SYSTEM: PC-DOS/MS-DOS
9      SOFTWARE: Patent In Release #1.0, Version #1.25
10
11     CURRENT APPLICATION DATA:
12
13     APPLICATION NUMBER: US/07/757,342D
14     FILING DATE: 10-Sep-1991
15
16     CLASSIFICATION: <unknown>
17
18     ATTORNEY/AGENT INFORMATION:
19
20     NAME: BUCKLEY, Linda M.
21     REGISTRATION NUMBER: 31003
22     REFERENCE/DOCKET NUMBER: 411226
23
24     TELECOMMUNICATION INFORMATION:
25
26     TELEPHONE: (617)523-3400
27     TELEFAX: (617)523-6440
28     TELEX: 200291 STRE UR
29
30     INFORMATION FOR SEQ ID NO: 6:
31
32     SEQUENCE CHARACTERISTICS:
33
34     LENGTH: 692 amino acids
35     TYPE: amino acid
36     TOPOLOGY: linear
37
38     MOLECULE TYPE: protein
39
40     SEQUENCE DESCRIPTION: SEQ ID NO: 6:
41
42     US-07-757-342D-6

```

```

Query Match 3.7%; Score 93; DB 4; Length 692;
Best Local Similarity 17.0%; Pred. No. 0.68;
Matches 83; Conservative 69; Mismatches 142; Indels 194; Gaps 20.

QY      79 HEDALTKTNILFVAIHREHYTSLMDLRHLVY-----KILIDVSNMR 121
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db      98 HEIRIEKANNLT-LYINPAFONLPBLRYLLTISNTGIKHLPAVHKIQSLQKVLIDIODIN 156
QY     122 INQYESMAEYIASLFPQSLIVKGGNVYSAMALQIGPKDASQVYICSNIIARQVIEL 181
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     157 IH-----IARNSEFGLSESYLWL-----SKNGIEIHNCA----- 189
QY     182 AROLNFIPIDGLSSAREIENLPRLFTLMGPPVVAISLATFEFLYSFVBDVIHPYAR 241
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     190 ---FNGTOLDELNLSDNNNLEELPNDVQGAGCVILDIS-----RTKVHSLPN 235
QY     242 NQOQDFYVPIPIEIVN-KLP---IVAITLISLVY-----LAGLAAAYOLY----- 283
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     236 HGLENLKRLRAKSTYRLKRLPNLDKQVYLMESAISLYPSHCACAFANLKROISELHPICKS 295
QY     284 -----YG-----TKYRRF-----PWLFTWLQOC 301
Db     296 ILRGDIDMTQIGDQVSLIDDEPYGKGSDMYNEFDIDLCNEVDVYCSKPPAFNFC 355
QY     302 RKQIG-----LISFFPAVHYA-----YSLCLPMKRSEERYELTMAYQOVHA 343
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     356 EDIMGYNLRVLINPISILATIGNTVLVLTTSQYKLLVP-----RFLMCHLAFAD--- 407
QY     344 NIENSNNEEYWRIMYISFGIMSGLISLAVTSIPSVSNLNMRE-----FSPI 394
Db     408 -----LCIGIYLLILIASVDIHITSQYHNHVAIDMQGTAGCAAGCEFFYF 450
QY     395 QSTLIG---YVALLISTFH-----VLIVGKRAFEERYRYFTTPNF 432
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     451 ASELSEYVTLTAITTERWHTITHAMQLECKVQLRHNAASVYVLCWTFAPAAALPPIFCISSY 510
QY     433 V-LALVP 439
      11 : : : : : : : : : : : : : : : : : : : : : : : :
Db     511 MKVSICLP 518

RESULT 14
US-07-841-997A-2
: Sequence 2, Application US/07841997A
: Patent No. 5422254

```





```

Db      80 KDLLEKFNVAVPFLSDEIADLHNGFNSNLSLWFLFHYHGEINF--ENAMG-YNEANQ 136
QY      131 -----EYLASLPDSLYKGF-----NVSAMALOLGPKDASROY 166
Db      137 TETNEIAKTMNNDLIWHDYHLMVPEMLRVKIHKKOLONKVGWFLH--PPPSSEIY 194
QY      167 ICSNNIOARQOYIE-----LAROLF----- 187
Db      195 ----RILPVROEILKGYLSCDVLGFEHTYDARHFLSSVQRYLVNNTLPNGVEYOGRFVNG 251
QY      188 ---IPIDLGSSARE-----IENLPLRLFTL---- 211
Db      252 AEPIDVDKFTDGLKKEGVQKRIQOLKFTPKCKIIVGVDRUDYIKGVQKRLHAMEVEL 311
QY      212 ----WRGPVY--VAISLATFFELYSFVADVIHPIARNOQSDPK---IPLEIVNKTLP 260
Db      312 NEHPERGRGVVAVQVAVPSRGDVEEYOYLRVSVNELVGRINGOFGTVEFVPIHFHKSIP 371
QY      261 IVAITLLSVYLAGLLAAAYQLYYGTKYRRFPPELFTWLQCRKOLGLSFFFAMVHAYS 320
Db      372 FEELISLVAUSDVCLVST-----RDGMNLVSY-----EYI 402
QY      321 LCLPMRSEERYL--FLNMAYQOVHANIENSWNEEYVRIEMTISFGIMSLGLSLAYTS 378
Db      403 ACOEKKGSLILSEFTGAQSLNGALITVNPWNTDLS-----AINEALT 447
QY      379 IPSVSNALWREPSFIQSTLGYVALLISTFHVILITGMKRAFEERY 424
Db      448 LPDVKKENVW-----EKLYIKSYTSAF-----WGENFVHELY 481

```

Search completed: May 9, 2002, 03:15:32  
 Job time: 11551 sec

THIS PAGE BLANK (USPTO)



PD 04-OCT-2001.  
 XX 23-MAR-2001; 2001WO-US09410.  
 XX 24-MAR-2000; 2000US-191929P.  
 XX (SAAT/) SAATCIOGLU F.  
 PA Saatioglu F.  
 PI  
 XX WPI: 2001-662926/76.  
 DR N-PSDB; AAS15793, AAS15801, AAS15802.  
 XX  
 PT New polynucleotide for the diagnosis, prevention and treatment for  
 PT prostate and testis disorders, particularly prostate cancer, comprises  
 PT prostate-specific or testis-specific nucleic acids  
 XX  
 PS Claim 1; Fig 4B; 114pp; English.  
 XX  
 CC The invention relates to substantially pure prostate-specific or  
 CC testis-specific polypeptides and the nucleic acids encoding them.  
 CC Also included are vectors and host cells expressing the proteins, a  
 CC transgenic animal expressing the protein, antibodies against the  
 CC proteins, probes for detecting the nucleic acids, antisense molecules  
 CC for the nucleic acids and methods of isolating modulators of the  
 CC proteins. Compounds that modulate the prostate specific or testis  
 CC specific polypeptide are useful to diagnose, prevent or treat disorders  
 CC of the testis or prostate particularly prostate cancer, benign  
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,  
 CC cryptorchidism, undescended, retractile, ascending or vanished  
 CC testis. Other proliferative disorders for which the modulators may be  
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast  
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The  
 CC present sequence represents a prostate specific protein, Six-  
 CC Transmembrane Protein of Prostate 1, STM1.  
 XX  
 SQ Sequence 490 AA;  
 Query Match 99.8%; Score 2530; DB 22; Length 490;  
 Best Local Similarity 99.8%; Pred. No. 4.8e-259;  
 Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MESISMNGSPKSLSECLPNCINGINKDARKVTGVIGSGDFAKSLTRLRCGYHVVIGS 60  
 DB 1 mesismngspkslsetclpncinginkdarkvtgvigsgdfaksltrlrcgyhvvigs 60  
 QY 61 RNPFASEFPFHVVDYHHEDEATKTNITFAIHREHTSLMDLRHLVGLIDVSNMM 120  
 DB 61 rnpfasefphvvdvthbedalktllfvalhrehyslwdlrhlhvgkllldvsnmm 120  
 QY 121 RINQYSPNSAEYLASFPPDSLIVKGFNVVSAMALQSGPKDASROYVICSNNIQAROOVIE 180  
 DB 121 rinqyspnsaeylaslfpdslivkgfnvvsamalgpkdasroyvicsnniqarqvie 180  
 QY 181 LARQNLNPIIDGLSSASAREIENLPLRLFTLMRGPVVVAISLAFFLYSFVRDVIHPYA 240  
 DB 181 larqnlndpiddglssasareienplrlftlmgpvvvaistlaafflysfvrdviphya 240  
 QY 241 RNOQSDRYKPIEIVNKTPIVATITLSLVYLAGLAAAYLYGTRRRPPPLEETWLO 300  
 DB 241 rnoqsdrykpielvnktlpivatitllsvylagllaaylygtrrrpppleetwldq 300  
 QY 301 CRKOLGLISFEFAWVAVSYSLCPMRRSERYLFNMAVYOOVHANINSMNEEVRITEMY 360  
 DB 301 crkqlglisffefamvavayslcpmrtserylfnnmayqyvhanienmneeevritemy 360  
 QY 361 ISFGIMSLGLSLLAVTSIDSVSNALNMWRRESFTQSTLGVYALLISFHVLLIYGMKRAFE 420  
 DB 361 isfgimslglslavtsidsvsnalnmwrresftqstlgyvallisfhnvliygwkrafe 420  
 QY 421 EEYRRETFPRNFVALVLPSTVILGKTIPLPCISRKLRKRKKWESQFLBESIGTIP 480  
 DB 421 eeyrretfprnfvalvlpstvilgktilplpcisrklrkrkkwesqflbesigtip 480

DB 421 eeyrretfprnfvalvlpstvilgktilplpcisrklrkrkkwesqfliegigtip 480  
 QY 481 HVSPEERYTVM 490  
 DB 481 hvspertyvm 490  
 RESULT 2  
 ID AAU10188 standard; Protein: 454 AA.  
 XX  
 AC AAU10188;  
 XX  
 XX 16-JAN-2002 (first entry)  
 DE Human ORF2 of Six-Transmembrane Protein of Prostate 1.  
 XX  
 KW Human; Six-Transmembrane Protein of Prostate 1; STM1; prostate cancer;  
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;  
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;  
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;  
 KW liver cancer; lung cancer; cytostatic; ORF2.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200172962-A2.  
 XX  
 XX 04-OCT-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09410.  
 PR  
 XX 24-MAR-2000; 2000US-191929P.  
 PA (SAAT/) SAATCIOGLU F.  
 XX  
 PI Saatioglu F.  
 XX  
 DR WPI: 2001-662926/76.  
 DR N-PSDB; AAS15810, AAS15811.  
 XX  
 PT New polynucleotide for the diagnosis, prevention and treatment for  
 PT prostate and testis disorders, particularly prostate cancer, comprises  
 PT prostate-specific or testis-specific nucleic acids  
 XX  
 PS Claim 1; Fig 4H; 114pp; English.  
 XX  
 CC The invention relates to substantially pure prostate-specific or  
 CC testis-specific polypeptides and the nucleic acids encoding them.  
 CC Also included are vectors and host cells expressing the proteins, a  
 CC transgenic animal expressing the protein, antibodies against the  
 CC proteins, probes for detecting the nucleic acids, antisense molecules  
 CC for the nucleic acids and methods of isolating modulators of the  
 CC proteins. Compounds that modulate the prostate specific or testis  
 CC specific polypeptide are useful to diagnose, prevent or treat disorders  
 CC of the testis or prostate particularly prostate cancer, benign  
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,  
 CC cryptorchidism, undescended, retractile, ascending or vanished  
 CC testis. Other proliferative disorders for which the modulators may be  
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast  
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The  
 CC present sequence is prostate specific protein, Six-Transmembrane  
 CC Protein of Prostate 1, STM1, ORF2.  
 XX  
 SQ Sequence 454 AA;  
 Query Match 90.5%; Score 2294; DB 22; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-234;  
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MESISMNGSPKSLSECLPNCINGINKDARKVTGVIGSGDFAKSLTRLRCGYHVVIGS 60  
 DB 1 mesismngspkslsetclpncinginkdarkvtgvigsgdfaksltrlrcgyhvvigs 60

OY 61 RNPFASEFPVVDVTHHEDALTKNIIFVAIHREYTSLMDLRHLVGLKILIDVSNM 120  
 |||||||  
 DB 61 rnpkfasefphvvdvthheda|tkn|l|fva|ihre|yts|lmd|lrhl|vgk|il|dvs|nm 120  
 |||||||  
 OY 121 RINQYPSNAEYLASLFPDSLIVKGFNVYSAMALQGPDKASROYICSNNIQARQOVI 180  
 |||||||  
 DB 121 ringypsnaey|las|l|f|ps|l|v|k|g|f|nv|ysa|ma|l|q|p|d|ka|s|ro|y|i|c|s|n|n|i|q|a|r|q|o|v|i 180  
 |||||||  
 OY 181 LARQNLFIPIIDIGLSSASAEIENLPLRLTLMRGVYVAISLATEFFLYSPVADVHPYA 240  
 |||||||  
 DB 181 larqnlfi|p|i|d|i|g|l|s|s|a|s|a|e|i|e|n|l|p|l|r|l|t|m|r|g|v|y|v|a|i|s|l|a|t|e|f|f|l|y|s|p|v|a|d|v|h|p|y|a 240  
 |||||||  
 OY 241 RNOQSDFYKPIEIVNKTPIVAITLSTLVYLAGLAAAYOYYGKRRPFWLTWTQ 300  
 |||||||  
 DB 241 rnoqsd|f|y|k|p|i|e|iv|n|k|t|p|i|v|a|i|t|l|s|t|l|v|y|l|a|g|l|a|a|y|o|y|y|g|k|r|r|p|f|w|l|t|w|t|q 300  
 |||||||  
 OY 301 CRKQGLSLSPFAMVAVSLCLPMRBSRYFLNMAVOOVANIENTSNNEEVRRIEMX 360  
 |||||||  
 DB 301 crkqgl|s|l|s|p|f|a|m|v|a|v|s|l|c|l|p|m|r|b|s|r|y|f|l|n|ma|v|o|o|v|a|n|i|e|n|t|s|n|n|e|e|v|r|r|i|e|m|x 360  
 |||||||  
 OY 361 ISFGIMSLGLSLAVTSIPSVSNALNMRSEFSTQSTGLCYALLSTFHVLIYGMKRAFE 420  
 |||||||  
 DB 361 isfgim|s|l|g|l|s|l|a|v|t|s|i|p|s|v|s|n|a|l|n|m|r|s|e|f|s|t|q|s|t|g|l|c|y|a|l|l|s|t|f|h|v|l|i|y|g|m|k|r|a|f|e 420  
 |||||||  
 OY 421 EEEYRFPNPFVIALVPSIVIL 444  
 |||||||  
 DB 421 eeyr|f|p|n|p|f|v|i|a|l|v|p|s|i|v|i|l 444  
 |||||||

RESULT 3  
 AAE02781  
 ID AAE02781 standard: Protein; 454 AA.

XX AAE02781;

DT 06-AUG-2001 (first entry)

XX Human six transmembrane epithelial antigen of prostate (STEAP)-2 protein.

XX Human; cytostatic; antiproliferative; vaccine; gene therapy;

KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;

KM chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;

KN pancreatic.

XX Homo sapiens.

XX Key

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Domain

Location/Qualifiers  
 100..108  
 /label= HLA-A2\_binding\_peptide #5  
 153..165  
 /label= Immunogenic\_peptide #1  
 207..228  
 /label= Transmembrane\_domain #1  
 227..235  
 /label= HLA-A2\_binding\_peptide #1  
 255..277  
 /label= Transmembrane\_domain #2  
 304..325  
 /label= Transmembrane\_domain #3  
 345..358  
 /label= Immunogenic\_peptide #2  
 359..381  
 /label= Transmembrane\_domain #4  
 393..415  
 /label= Transmembrane\_domain #5  
 402..410  
 /label= HLA-A2\_binding\_peptide #2  
 428..450

FT XX /label= Transmembrane\_domain #6  
 PN XX WO200140276-A2.  
 PD XX 07-JUN-2001.  
 PF XX 06-DEC-2000; 2000WO-US33040.  
 PK XX 06-DEC-1999; 99US-0455486.  
 PA XX (UROG-) UROGENESYS INC.  
 PI PI Afar DEH, Hubert RS, Raitano AB, Safran DC, Mitchell SC, Faris M;  
 DR DR Jakobovits A;  
 DR DR WPI: 2001-367804/38.  
 DR DR N-PSDB; AAD07072.  
 XX XX  
 XX XX  
 PS PS Claim 1; Fig 9A-9C; 187pp; English.  
 CC CC The present sequence is human six transmembrane epithelial antigen of  
 CC CC the prostate (STEAP)-2 protein. STEAP is a member of cell  
 CC CC surface serpentine transmembrane antigens. STEAP-2 gene is located on  
 CC CC chromosome 7q21 and is used in gene therapy. Inhibiting the development  
 CC CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian  
 CC CC and pancreatic) expressing STEAP or inhibiting growth or killing cells  
 CC CC expressing STEAP in a patient, comprises administering a vaccine  
 CC CC composition to the patient. Treating a patient with a cancer that  
 CC CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,  
 CC CC comprises administering to the patient a vector encoding single chain  
 CC CC monoclonal antibody that comprises the variable domains of the heavy and  
 CC CC light chains of the monoclonal antibody that specifically binds to STEAP,  
 CC CC such that the vector delivers the single chain monoclonal antibody coding  
 CC CC sequence to the cancer cells and the encoded single chain monoclonal  
 CC CC antibody is expressed intracellularly.  
 CC CC Note: This sequence is stated to be the same as that being shown as  
 CC CC SEQ ID NO:8 (AAE02841) in sequence listing of the specification. However  
 CC CC both the sequences differ at several positions.  
 SQ XX Sequence 454 AA:  
 Query Match 90.5%; Score 2294; DB 22; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-234;  
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MESISMGSPKSLSECTCLNGINGRIDARKVYGVYIGSGDFAKSLTRILRCYRHVIGS 60  
 |||||||  
 DB 1 mesismgspk|s|l|s|e|c|t|c|l|n|g|i|n|g|i|r|i|d|a|r|k|v|y|g|v|y|i|g|s|g|d|f|a|k|s|l|t|r|i|l|r|c|y|r|h|v|i|g|s 60  
 |||||||  
 OY 61 RNPFASEFPVVDVTHHEDALTKNIIFVAIHREYTSLMDLRHLVGLKILIDVSNM 120  
 |||||||  
 DB 61 rnpkfasefphvvdvthheda|tkn|l|fva|ihre|yts|lmd|lrhl|vgk|il|dvs|nm 120  
 |||||||  
 OY 121 RINQYPSNAEYLASLFPDSLIVKGFNVYSAMALQGPDKASROYICSNNIQARQOVI 180  
 |||||||  
 DB 121 ringypsnaey|las|l|f|ps|l|v|k|g|f|nv|ysa|ma|l|q|p|d|ka|s|ro|y|i|c|s|n|n|i|q|a|r|q|o|v|i 180  
 |||||||  
 OY 181 LARQNLFIPIIDIGLSSASAEIENLPLRLTLMRGVYVAISLATEFFLYSPVADVHPYA 240  
 |||||||  
 DB 181 larqnlfi|p|i|d|i|g|l|s|s|a|s|a|e|i|e|n|l|p|l|r|l|t|m|r|g|v|y|v|a|i|s|l|a|t|e|f|f|l|y|s|p|v|a|d|v|h|p|y|a 240  
 |||||||  
 OY 241 RNOQSDFYKPIEIVNKTPIVAITLSTLVYLAGLAAAYOYYGKRRPFWLTWTQ 300  
 |||||||  
 DB 241 rnoqsd|f|y|k|p|i|e|iv|n|k|t|p|i|v|a|i|t|l|s|t|l|v|y|l|a|g|l|a|a|y|o|y|y|g|k|r|r|p|f|w|l|t|w|t|q 300  
 |||||||  
 OY 301 CRKQGLSLSPFAMVAVSLCLPMRBSRYFLNMAVOOVANIENTSNNEEVRRIEMX 360  
 |||||||  
 DB 301 crkqgl|s|l|s|p|f|a|m|v|a|v|s|l|c|l|p|m|r|b|s|r|y|f|l|n|ma|v|o|o|v|a|n|i|e|n|t|s|n|n|e|e|v|r|r|i|e|m|x 360  
 |||||||

QY 361 ISFGIMSLGLSLAIVTSIPSVSNALMWRESFSTQSTLGYVALLISFPHVLLYGMKRAFE 420  
DB 361 ISFGIMSLGLSLAIVTSIPSVSNALMWRESFSTQSTLGYVALLISFPHVLLYGMKRAFE 420  
QY 421 EEEYRFTTPPNFVVALVPSIVIL 444  
DB 421 eeyryfttppnfvalvpsivil 444

RESULT 4  
ID AAU10189 standard; Protein: 419 AA.  
AC AAU10189;  
DE 16-JAN-2002 (first entry)  
XX Human ORF3 of Six-Transmembrane Protein of Prostate 1.  
XX Human; Six-Transmembrane Protein of Prostate 1; STM1; prostate cancer;  
KM benign prostatic hyperplasia; acute prostatitis; testicular cancer;  
KM cryptorchidism; testicular disorder; proliferative disorder; lymphoma;  
KM leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;  
KM liver cancer; lung cancer; cytostatic; ORF3.  
XX Homo sapiens.  
XX WO200172962-A2.  
XX 04-OCT-2001.  
XX 23-MAR-2001; 2001WO-US09410.  
XX 24-MAR-2000; 2000US-191929P.  
XX (SAAT/) SAATCTOGLU F.  
XX Saactioglu F;  
XX WPI: 2001-662926/76.  
XX N-PSDB; AAS15813.  
XX New polynucleotide for the diagnosis, prevention and treatment for  
PT prostate and testis disorders, particularly prostate cancer, comprises  
PT prostate-specific or testis-specific nucleic acids -  
XX Claim 1; Fig 4K; 114pp; English.

XX The invention relates to substantially pure prostate-specific or  
CC testis-specific polypeptides and the nucleic acids encoding them.  
CC Also included are vectors and host cells expressing the proteins, a  
CC transgenic animal expressing the protein, antibodies against the  
CC proteins, probes for detecting the nucleic acids, antisense molecules  
CC for the nucleic acids and methods of isolating modulators of the  
CC proteins. Compounds that modulate the prostate specific or testis  
CC specific polypeptide are useful to diagnose, prevent or treat disorders  
CC of the testis or prostate particularly prostate cancer, benign  
CC prostatic hyperplasia, acute prostatitis, testicular cancer,  
CC cryptorchidism, undescended, retractile, ascending or vanished  
CC testis. Other proliferative disorders for which the modulators may be  
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast  
CC cancer, pancreatic cancer, liver cancer and lung cancer. The  
CC present sequence is prostate specific protein, Six-Transmembrane  
CC Protein of Prostate 1, STM1, ORF3.  
XX  
XX Sequence 419 AA;

Query Match 80.3%; Score 2036; DB 22; Length 419;  
Best Local Similarity 99.7%; Pred. No. 8,6e-207;  
Matches 394; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MESISMGSPKSLSEETCLPNCINGIKDARKVTGVCISGDFAKSLTIRLRGCVHVVIGS 60  
DB 1 mesismgspkslsetclpncingikdarkvtgvcisgdfaksltirrcgvhvvigs 60  
QY 61 RNPKEASEFPFHVVDYTHHEDALTKNTNIFVAIHREHTSLMDLRHLJYCKILLDVSNNM 120  
DB 61 rnpkasefphvvdvthhedalkntnifvaihrehtslmdlrhllyckilldvsnnm 120  
QY 121 RINQYPSNAEYSLAPFDSLIVKGFNVVSAMALQLOPKDASROYVICSNNIQAQOOVIE 180  
DB 121 ringypsnaeyslapfDSLIVKGFNVVSAMALQLOPKDASROYVICSNNIQAQOOVIE 180  
QY 181 LARQLNFPIDLGSLSSARETENLPLRLFTLMRGPVVAISLATFEFLYSFVROVHPYA 240  
DB 181 larqlnfpidlgslssaretenlplrlftlmrgpvvaisslatfeFLYSFVROVHPYA 240  
QY 241 RNOQSDPKPIPIETVNTKTPRIVATTLISIVYLAQLLAAQVLYGTVRRPPPLETWLQ 300  
DB 241 rnoqsdpykpiPIETVNTKTPRIVATTLISIVYLAQLLAAQVLYGTVRRPPPLETWLQ 300  
QY 301 CRKOLGLSPFAMVHVAYSICLPMRSELYFLNMAYOOVHANIEENSMNEEYVRIEMY 360  
DB 301 crkqlglspffamvhvaysiclpmrseLYFLNMAYOOVHANIEENSMNEEYVRIEMY 360  
QY 361 ISFGIMSLGLSLAIVTSIPSVSNALMWRESFSTQ 395  
DB 361 ISFGIMSLGLSLAIVTSIPSVSNALMWRESFSTQ 395

RESULT 5  
ID ABG12306 standard; Protein: 576 AA.  
XX ABG12306;  
XX ABG12306;  
XX 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #12297.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX N-PSDB; AAS76493.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID NO 42665; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 576 AA:

Query Match 77.4%; Score 1962.5; DB 22; Length 576;  
 - Best Local Similarity 87.4%; Pred. No. 8.6e-199;

Matches 388; Conservative 0; Mismatches 1; Indels 55; Gaps 1;

QY 1 MESISMGSPKLSSETCLPENGINGIDARKVTVGVIGSGDFAKSLTIRLCGYHVVIGS 60  
 DB 1 mesismgspklssetclpungingidarkvtvgvigsdfaksltirrcgyhvvigs 60  
 QY 61 RNPFASEFPFHVVDTHHEDALTKTNIIFFVAHREHYTSLMDRLYLKILIDVSNM 120  
 DB 61 rnpkfasefpfhwvdthhedaltknllfvalhrehytslwdrlhlygkllidvsnm 120  
 QY 121 RINQYPSNAEYLASLPDLSLVKGFNVVSAMALQGPDKASROYVICSNNIQAROOVIE 180  
 DB 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 QY 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 DB 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 QY 181 LARQLNFIPIIDGSLSSAREIENLPRLFTLMRGPPVVAISLATFFFLYSFVADVTHPYA 240  
 DB 181 larqlnfiipidgslssareienlprlftlwrgppvvaistlatffflysfvadvthpya 240  
 QY 241 RNOQSDFYKPIEIVNKTPIVAITLISLVYLAGLAAAYQLYGYKRRFPMLETWIQ 300  
 DB 241 rnoqsdfykpielvknktlpivaitllslvylagllaaayqlygykrrfpmletwlg 300  
 QY 301 CRKOLGLSFPFAMVAVAYSLCLPMRSERYFLNMAVOOVANINENSNNEEVRRIEMV 360  
 DB 301 crkqglglslffamvavayslclpmrserylflnmayq----- 339  
 QY 361 ISFGIMSLGLSLAVTSIPSVSNALNMRFESEFSTIGVALLISTFHVLYIGMKRAPE 420  
 DB 361 isfgimslglslavtspsvsnalnmrfeesefstigvallistfhvlyigmkrape 420  
 QY 361 isfgimslglslavtspsvsnalnmrfeesefstigvallistfhvlyigmkrape 420  
 DB 361 isfgimslglslavtspsvsnalnmrfeesefstigvallistfhvlyigmkrape 420  
 QY 421 EEEYRFTPTPNFVLAIVLPSTIVL 444  
 DB 421 eeyrftptpnfvlavlpstivl 389  
 QY 421 eeyrftptpnfvlavlpstivl 389  
 DB 421 eeyrftptpnfvlavlpstivl 389

RESULT 6

ID ABG00113 standard; Protein; 1273 AA.

AC ABG00113;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #104.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR N-PSDB; AAS64300.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

Claim 20; SEQ ID No 30472; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Seq Sequence 1273 AA;

Query Match 77.4%; Score 1962.5; DB 22; Length 1273;

Best Local Similarity 87.4%; Pred. No. 2.9e-198;

Matches 388; Conservative 0; Mismatches 1; Indels 55; Gaps 1;

QY 1 MESISMGSPKLSSETCLPENGINGIDARKVTVGVIGSGDFAKSLTIRLCGYHVVIGS 60  
 DB 1 mesismgspklssetclpungingidarkvtvgvigsdfaksltirrcgyhvvigs 60  
 QY 61 RNPFASEFPFHVVDTHHEDALTKTNIIFFVAHREHYTSLMDRLYLKILIDVSNM 120  
 DB 61 rnpkfasefpfhwvdthhedaltknllfvalhrehytslwdrlhlygkllidvsnm 120  
 QY 121 RINQYPSNAEYLASLPDLSLVKGFNVVSAMALQGPDKASROYVICSNNIQAROOVIE 180  
 DB 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 QY 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 DB 121 rinqypsnaeylaslfpdlsllvkgnvvsawaldqpkdasroyvicsnniqarqvle 180  
 QY 181 LARQLNFIPIIDGSLSSAREIENLPRLFTLMRGPPVVAISLATFFFLYSFVADVTHPYA 240  
 DB 181 larqlnfiipidgslssareienlprlftlwrgppvvaistlatffflysfvadvthpya 240  
 QY 241 RNOQSDFYKPIEIVNKTPIVAITLISLVYLAGLAAAYQLYGYKRRFPMLETWIQ 300  
 DB 241 rnoqsdfykpielvknktlpivaitllslvylagllaaayqlygykrrfpmletwlg 300  
 QY 301 CRKOLGLSFPFAMVAVAYSLCLPMRSERYFLNMAVOOVANINENSNNEEVRRIEMV 360  
 DB 301 crkqglglslffamvavayslclpmrserylflnmayq----- 411

QY 361 ISFIMSLGLSLAVTSIPSVNALNMRFFSTIOSTLGVALLSTFHVLIYGMKRAFE 420  
 DB 412 -----gstlgyalliststhvliygwkrafe 437  
 QY 421 EEYRFTYPPNFVALVLPSTIVIL 444  
 DB 438 eeyrftypnftvalvlpstivil 461  
 RESULT 7  
 ID AAE02841 standard; protein; 450 AA.  
 AC AAE02841;  
 DT 06-AUG-2001 (first entry)  
 DE Human STEAP-2 protein, alternative version.  
 XX  
 KM Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KM six transmembrane epithelial antigen of the prostate-2; STEAP-2;  
 KM chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;  
 KM pancreatic.  
 XX  
 OS Homo sapiens.  
 XX  
 Key location/Qualifiers  
 FH Misc-difference 335..336  
 FT Misc-difference /note="Encoded by GACTGAGCT"  
 FT Misc-difference 375..376  
 FT Misc-difference /note="Encoded by AGATGAAGT"  
 FT Misc-difference 415..416  
 FT Misc-difference /note="Encoded by GAGTAAGC"  
 FT Misc-difference 445..446  
 FT Misc-difference /note="Encoded by ACATGAAGT"  
 FT Misc-difference 448..449  
 FT Misc-difference /note="Encoded by AATTAAATTC"  
 XX  
 PN WO200140276-A2.  
 PD 07-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 PA (UROC-) UROGENESYS INC.  
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Paris M;  
 PI Jakobovits A;  
 XX  
 DR WPI; 2001-367804/38.  
 DR N-PSDB; AAD07072.  
 XX  
 PT New STEAP (six transmembrane epithelial antigen of the prostate)  
 PT proteins, expressed in human cancers, useful for detecting and treating  
 PT cancer.  
 XX  
 PS Claim 1; Page 175-176; 187pp; English.  
 XX  
 CC The present sequence is an alternative version of human six transmembrane  
 CC epithelial antigen of the prostate (STEAP)-2 protein. STEAP is a member  
 CC of cell surface serpentine transmembrane antigens. STEAP-2 gene is  
 CC located on chromosome 7q21 and is used in gene therapy. Inhibiting the  
 CC development or progression of a cancer (eg. prostate, colon, bladder,  
 CC lung, ovarian and pancreatic) expressing STEAP or inhibiting growth or  
 CC killing cells expressing STEAP in a patient, comprises administering a  
 CC vaccine composition to the patient. Treating a patient with a cancer that  
 CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,  
 CC comprises administering to the patient a vector encoding single chain  
 CC monoclonal antibody that comprises the variable domains of the heavy and  
 CC light chains of the monoclonal antibody that specifically binds to STEAP.

CC such that the vector delivers the single chain monoclonal antibody coding  
 CC sequence to the cancer cells and the encoded single chain monoclonal  
 CC antibody is expressed intracellularly.  
 CC Note: This sequence is stated to be the same as that being shown as  
 CC SEQ ID NO:8 (AAE02781) in figure 9A-9C of the specification. However  
 CC the present sequence lacks several amino acids at its N-terminal end  
 CC and has additional amino acids at its C-terminal end when compared with  
 CC the sequence shown in figure 9A-9C.  
 XX  
 SQ Sequence 450 AA;  
 Query Match 66.4%; Score 1683; DB 22; Length 450;  
 Best Local Similarity 94.6%; Pred. No. 2.5e-169;  
 Matches 332; Conservative 1; Mismatches 8; Indels 10; Gaps 1;  
 QY 119 NMRINQPSNAEYLAIFPDSLVKGFNVYSANALQGPDAARQYICNNIQAAQOV 178  
 DB 1 nmrlndqpsnaeylaifpdsllvkgfnvsawalqlpksarqyicnnlqarqv 60  
 QY 179 IELAROLNFTPIDGLSSAREIENLPLRLFTLMRGPVVAISLATEFFELXSFVRYIHP 238  
 DB 61 Ielarglnflpidglssareienlplrlftlwrpvyvaaislatffllysfvrydihp 120  
 QY 239 YARNQSDFYKPIPIEYVNTLPPIVATLTLVYLAGLAAAYQLYGTXYRRPPELWTW 298  
 DB 121 yarngsdffykipielpienvktlppivaatlslvylagllaayqlygtkyrrfpplwtw 180  
 QY 299 LQCRKQGLSFFRPMHVAISLCLPQRSRYFLFMNAQOVHANNENSGNEEVRIE 358  
 DB 181 lqcrkqglslsfframhvayslclpqrserylflmayqyvhaanlensvneevrle 240  
 QY 359 MYISFGIMSLGLSLAVTSIPSVNALNMRFFSTIOSTLGVALLSTFHVLIYGMKRA 418  
 DB 241 myisfgimslglslavtsipsvnalnmrffstlsglyvallystfhlvlygkra 300  
 QY 419 FEEERYRFTYPPNFVALVLPSTIVILGKIIL-----FLPCISRKRL 459  
 DB 301 feeeryrftypnftvalvlpstivildlqlcrypdagtgiclpdstslk 351  
 RESULT 8  
 ID AAB49483 standard; protein; 488 AA.  
 AC AAB49483;  
 DT 08-MAR-2001 (first entry)  
 DE Rat P-HYDE.  
 XX  
 KM Rat; P-HYDE; cytostatic; gene therapy; apoptosis; leukemia; prostate;  
 KM tumour suppressor gene; DNA repair; cancer; melanoma; lymphoma;  
 KM colorectal; pancreatic; breast; brain; gastric carcinoma.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200071564-A2.  
 PD 30-NOV-2000.  
 XX  
 PF 01-MAY-2000; 2000WO-US11456.  
 XX  
 PR 29-APR-1999; 99US-0131607.  
 PR 29-APR-1999; 99US-0302457.  
 PR 26-NOV-1999; 99US-0499817.  
 XX  
 PA (UYTE-) UNIV TENNESSEE RES CORP.  
 PI Steiner MS, Wang C, Rinaldy A, Menon R;  
 XX  
 DR WPI; 2001-032016/04.  
 DR N-PSDB; AAC89167.





```

Db      477 tchalaekts 486
|:
RESULT 10
AAU10220
ID      AAU10220 standard; Protein; 488 AA.
XX
AC      AAU10220;
DT      16-JAN-2002 (first entry)
DE      Human Six-Transmembrane Protein of Prostate 3, STMP3.
XX
KW      Human; Six-Transmembrane Protein of Prostate 3; STMP3; prostate cancer;
KW      benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW      cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW      leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW      liver cancer; lung cancer; cystostatic.
XX
OS      Homo sapiens.
PN      WO200172962-A2.
PD      04-OCT-2001.
XX
PR      23-MAR-2001; 2001WO-US09410.
XX
PR      24-MAR-2000; 2000US-191929P.
PA      (SAAT/) SAATC10GLU F.
XX
PI      Saatic10glu F;
XX
DR      WPI; 2001-662926/76.
DR      N-PSDB; AAS15815.
XX
PT      New polynucleotide for the diagnosis, prevention and treatment for
PT      prostate and testis disorders, particularly prostate cancer, comprises
PT      prostate-specific or testis-specific nucleic acids
XX
PS      Claim 1; Fig 4M; 114pp; English.
XX
CC      The invention relates to substantially pure prostate-specific or
CC      testis-specific polypeptides and the nucleic acids encoding them.
CC      Also included are vectors and host cells expressing the proteins, a
CC      transgenic animal expressing the protein, antibodies against the
CC      proteins, probes for detecting the nucleic acids, antisense molecules
CC      for the nucleic acids and methods of isolating modulators of the
CC      proteins. Compounds that modulate the prostate specific or testis
CC      specific polypeptide are useful to diagnose, prevent or treat disorders
CC      of the testis or prostate particularly prostate cancer, benign
CC      prostatic hyperplasia, acute prostatitis, testicular cancer,
CC      cryptorchidism, undescended, retractile, ascending or vanished
CC      testis; Other proliferative disorders for which the modulators may be
CC      used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC      cancer, pancreatic cancer, liver cancer and lung cancer. The
CC      present sequence is a prostate specific protein, Six-Transmembrane
CC      Protein of Prostate 3, STMP3.
XX
SQ      Sequence 488 AA;
|:
Query Match 53.5%; Score 1356; DB 22; Length 488;
Best Local Similarity 53.1%; Pred. No. 1,3e-134;
Matches 260; Conservative 91; Mismatches 121; Indels 18; Gaps 6
|:
3 SISMGSPKSLSETCLPNCINGINKDKARKYTVGVSSGDFAKSLFTLRIGYHVYIGSRN 62
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
11 shldssslak--vpd-----aap--vgllsgsfarslatrlvsgfkvvvgsrn 60
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
63 PKFASEFRPHVVDYTHNEDALFKTNIFPAHIRENTYSIMDLRHLVGLILDVSN--NM 120
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

```

Db	61	pkrtarilypsaagvtcfgeaavsspevifvafrehyslscslsdqagkllvdvsnpteq	120
Qy	121	RINQPPSNAETYLASLPDLSLYKGFENVVSAMALQIGPKDASQOYICSNINIAQOQVIE	180
Db	121	ehlgqreneaeylasilfpctcvvafnvlavtqdgqprgdnqyplvcgdqpeakravse	180
Qy	181	LAROLNFPIDIGSSSAREIENPLEFLFTLMRGQPVVAVTSLATPEFLYSFVRDVIHPA	240
Db	181	malangrmpvimgslaaaweeampitllpwkvpkrlialgltvcfyaynfvrdrvlpqv	240
Qy	241	RMOQSDFPYKIDIEIVNKTLPVATITLLSLVYLAGLLAAAYQLYGTRKRRFPWLETWQ	300
Db	241	gesqmkfeklpsvsvntllpcvayvllslvyllpgvlaaaqlrrgkkyqfpoawlghwiq	300
Qy	301	CRKQGLLSFFFAWHVAVSYCLPEMRSEERYLFELNMAQOQVHANINSNNEEPVMEIYK	360
Db	301	hrtqgllstfccaahalysctclprrahrydvlvnlavkvlanckhllweewetmely	360
Qy	361	ISFGIMSLGSLSLAVTISPEVSNALMWRREFSIQSTLGVALLISTPHVLYIGMKRAFE	420
Db	361	lslvgialgtlslslavslpsianslmwrefsfqsslgvialvslhtllygyvrrate	420
Qy	421	EERYEYTPPNEFVALVPSIVILGKILLEPCISRKRLKIKGWEKSOFLBEGIGTIP	480
Db	421	esrykfyfppfcitlllypcvllakallfpccsrllarlrtrgwer-----estlkftlp	476
Qy	481	--HVSPEYVT 488	
Db	477	tdhalaekts 486	
RESULT 11			
ID	AAU04564	standard; Protein; 488 AA.	
XX	AAU04564;		
XX	17-DEC-2001	(first entry)	
XX			
DE		Human G-protein coupled receptor like protein, GPCR #4.	
KW		Human; G-protein coupled receptor like protein; GPCR; immunogen;	
KW		ophthalmic disease; neurological disease; Alzheimer's disease;	
KW		Parkinson's disease; immunological disorder; HIV; candidiasis;	
KW		human immunodeficiency virus; autoimmune disorder; multiple sclerosis	
KW		systemic lupus erythematosus; rheumatoid arthritis; platelet disorder	
KW		thrombocytopaenia; aplastic anaemia; inflammatory disorder;	
KW		septic shock; systemic inflammatory response syndrome; SIRS;	
KW		hormonal dysfunction; cancer; atherosclerosis; wound; STRAP-1;	
KW		tissue regeneration; haemophilia; leukaemia; reperfusion injury;	
KW		psoriasis; diabetes.	
OS		Homo sapiens.	
XX			
XX		WO200153454-A2.	
XX			
PD		26-JUL-2001.	
XX			
PF		22-DEC-2000; 2000MO-US34983.	
XX			
PR		21-JAN-2000; 2000US-0488725.	
PR		25-APR-2000; 2000US-0592317.	
PR		20-JUN-2000; 2000US-0598042.	
PR		19-JUL-2000; 2000US-0620312.	
PR		31-AUG-2000; 2000US-0653450.	
PR		04-DEC-2000; 2000US-0729739.	
XX			
PA		(HYSEQ-) HYSEQ INC.	
XX			
PI		Yamazaki V, Tang YT, Liu C, Zhou P, Wang D, Zhang J, Ren F;	
PI		Asundi V, Drmanac RT.	
XX			
DR		WPI: 2001-442255/47.	

DR N-PSDB; AAS08646.

PT New G-protein-coupled receptor-like polypeptides and polynucleotides, useful for treating diseases of ophthalmic, neurological, immunological and nephritic systems and hormonal dysfunction, cancer, atherosclerosis and diabetes -

PS Claim 10; Page 164-166; 259pp; English.

CC The sequence represents a human G-protein coupled receptor (GPCR)-like  
CC protein, found to have sequence similarity to human STRAP-1 protein.  
CC The GPCR-like polypeptides and polynucleotides are useful for  
CC the treatment of diseases of ophthalmic, neurological (e.g. Alzheimer's  
CC disease and Parkinson's disease, immunological (e.g. HIV infection and  
CC candidiasis), autoimmune disorders (e.g. multiple sclerosis, systemic  
CC lupus erythematosus and rheumatoid arthritis), platelet disorders (e.g.  
CC thrombocytopaenia and aplastic anaemia), inflammatory disorders (e.g.  
CC septic shock and systemic inflammatory response syndrome, SIRS) and  
CC nephritic systems. They may also be used to treat hormonal dysfunction,  
CC cancer, atherosclerosis, wounds, tissue regeneration, hemophilia,  
CC leukemias, reperfusion injury, psoriasis and diabetes. Numerous examples  
CC of each type of disorder are given in the specification. Anti-GPCR-like  
CC protein antibodies are useful for detecting or quantitating the  
CC polypeptide in tissue. The polypeptides can also be used as molecular  
CC weight markers and as a food supplement.

SQ Sequence 488 AA;

Query Match	53.5%;	Score 1356;	DB 22;	Length 488;
Best Local Similarity	53.18;	Pred. No. 1.3e134;		
Matches 260; Conservative	91;	Mismatches 121;	Indels 18;	Gaps 6;

```

0Y 3 SIEMASPSISITCPLPINCINGIKARKVWVGVGSDGFKSLITLIRGCHVYVGSN 62
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 11 slhvwdsdslaK--vvd-----eaK--vgllgsdfratslrlvsglKvvgvgrn 60
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 63 PKFASEFPBVVDVTHBEDALTKTILFEVALIHREHYTSLMDLHNLVGRKILDIVSN--NM 122
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 61 pKtatlrysaagvlfgeaavspvllfvafrfhyslcsldqIagklIvdvsnpteq 122
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 121 RINQYESNAKEYLASLPEPDSLVKGFNVSAMALOLGPKRASROYIGSNINQARQYIE 18
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 121 ehlgfhresnaeaylaaflfrctcvyakaftmawlsawclgaqprlgmqvdrldqgpaKraVse 18
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 181 LAROLNFIDIDGSLSSAREIEINLEPLRFTLMGRPVVAALSLATFPEFLYSFVADYIH9YA 24
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 181 maIamgfmpvmdngsIasaeveamplrlrlpwkvyrtllaIglfvcfeyaynfrvdrIqrvy 24
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 241 RNOQSOFFYKLEIYNKTLPIYATILTSIYVGLGSLAAAOVLGKRYRRFPYMETMQ 30
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 241 qesqnfKfKlpayvvnctIpcvayylslvyIpgvIaaalqlrtgkkyqfdrwIdmhwI 30
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 301 CRKQGLGLSFFPAMVHVAVSLCLPMRSEERYLFPMNAVQOVHANINMSNMNEEVRMIEMX 36
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 301 hrqqlglslffcaalhalysfclprtrahydlvnlavkvgIankshlIwveevwmIely 36
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 361 ISGCSIGLSLSLAWTSIPVSANLMMNRRSESFQSLGVLGVALLISTFPHVLIYGKRAPE 42
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 361 lslgvalaIclslslavtcsplstanslwnreIstysqslglvalavlsIthltfygwtrafe 42
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 421 EBYVRFYPPNPNVLAIVLPSYIILGKILFLPCISRLKRLIKKKGWESOFLEEGJAGTGP 48
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 421 esryKfYlppftclIvlpvcvllakalflIpcIstrlarlrrgwer-----estlkfclp 47
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

0Y 481 --HVSPEERYT 488
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 477 tchnalaekts 486
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

RESULT	12
AAB93224	
ID	AAB93224 standard; Protein; 488 AA

XX AAB93224;  
AC

DT 26-JUN-2001 (first entry)

DE	Human protein sequence SEQ ID NO:12212.
DE	Human protein sequence SEQ ID NO:12212.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 11-JAN-2000; 2000JP-0118776.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa

XX XX

XX

PT full-length cDNAs defined in the specification, and for the detection

PT full-length cDNAs -

PS Claim 8; SEQ ID 12212; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

SQ Sequence 488 AA;

Query Match	53.5%	Score 1356;	DB 22;	Length 488;
Best Local Similarity	53.1%;	Pred. No. 1.3e-134;		
Matches 260; Conservative	91;	Mismatches 121;	Indels 18;	Gaps 6;

```
OY      3  SIMSMGSPKLSIEFCELENGICINDARKVYGVGSDDPAKSLTILRLRCGHVHTSSRN 62
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     11  slhlvdscslak--vdpd-----eapk--vgllsgsdffarslatrlvgsfgfkvvysgm 60
OY      63  PKFASEFPFHVVDTTHHEDALTKTNIIIFVALIHREHYTSLMDLRHLVLGKITLIDVSN--NM 120
```



```

XX 17-AUG-1999; 99US-0149641.
PR 09-NOV-1999; 99US-0164203.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Yue H, Tang YT, Bandman O, Burford N, Azimzai Y;
PI Baughn MR, Lu DM, Patterson C;
XX
DR WPI: 2001-168860/17.
DR N-PSDB: AAF81761.
XX
PT Isolated polypeptide with a human membrane associated protein sequence
PT is useful for the diagnosis, prevention and treatment of cell
PT proliferative, autoimmune/inflammatory, neurological and
PT gastrointestinal disorders -
XX
PS Claim 1; Page 134-135; 173pp; English.
XX
CC AAF81741 to AAF81777 encode the human membrane associated proteins
CC (MEMAP) given in AAB74695 to AAB74731. MEMAPs have cytostatic,
CC antiinflammatory, anticonvulsant, immunosuppressive, antidiabetic and
CC antiarteriosclerotic activities, which can be used in gene therapy.
CC MEMAPs and agonist of MEMAPs can be used to treat a disease or condition
CC associated with decreased expression of functional MEMAP and antagonists
CC of MEMAP are used to treat a disease or condition associated with
CC overexpression of functional MEMAP. These disorders include cell
CC proliferative, autoimmune/inflammatory, neurological and gastrointestinal
CC disorders. The MEMAP polynucleotides and proteins are also used for the
CC diagnosis of these disorders. Specific examples of these disorders
CC include cancer, inflammation, atherosclerosis, epilepsy and diarrhoea.
CC MEMAP proteins can be used to screen for compounds which specifically
CC bind MEMAP including antibodies, oligonucleotides, proteins and small
CC molecules. MEMAP polynucleotides can be used to prepare transgenic
CC animals which can be studied to provide information concerning human
CC disease. Anti-MEMAP antibodies are useful in immunoassays for the
CC detection of MEMAP protein and can be used as antagonists to treat or
CC prevent a disorder associated with MEMAP. Polynucleotides encoding MEMAP
CC can be delivered to target cells with genetic abnormalities with respect
CC to the expression of MEMAP to treat or prevent a disorder associated
CC with MEMAP.
XX
XX Sequence 459 AA:
SQ

```

```

Query Match 45.6%; Score 1156; DB 22; Length 459;
Best Local Similarity 46.6%; Pred. No. 1.8e-113;
Matches 214; Conservative 96; Mismatches 143; Indels 6; Gaps 3;

```

```

OY 13 LSETC--LPNGINGIKDRKRVVGVIGSGDFAKSLTIRLRCGYHVVIGSRMPKASEF 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 mectcdaipicm--ssekqetvcifgtdgfgfgrslgikmkgqyvvfvsrnpq-ktel 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 70 FPHVVDVTHHEDALTKNTNIFVAIRREHYTSLMDLRLHVLGKILIDVSNMRINQYPSN 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 lpsgaevlvsyseaakskdilliahrehydflltellevngklldvismnlknqpsen 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 130 AETLASLPDLSLVKGFNVVSAMALQGRKDSROYTCSNNIQAOQVIELARQLNFTIP 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 aeylahlvpghavkafntisawalysgaldsargyfvsgndskakqymdlyrnlgfip 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 190 IDGSSASAREIEMLPLRFTLGRGPVVAISLATPEFLYSFVDVYIHPARNOQSDFKK 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 mdqsglmaakelekypdqlgfmwtrfrylsavlcvfifcyvrdvlypyvekkndtfr 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 250 IPRIEYKTLPLVATITLSLVYLAGLLAAAYOLYGTGKRRFPWLETWLOCRQGLLS 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 malsipnrlfptaltllalalypryiaaillgyrtktyrrfpdwdlhmmlcrkqglva 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 310 FFFAMVHVAVSLCPMRBSERYFLMAYAOQVANIENSMNEEVRRIEYIFGIMSLG 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 298 lqfatihvlytlvrltyvvtwrlgnlvtvgaillkknepfstssawlsdysvalgillgff 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 370 LLSLAVTSIPSVSNALMNRFPISQSTGLVALLSTFHVLLYGMKRAFEERYREYTP 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 358 lfvllgitspsvsnvneffvysklyglillctaltlllyggkfrlspnltwylp 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 430 PNFVALVLPSTVILGKIILFLPCISRKLRKIKKGEKS 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 418 aayvlglilipctvilkfvlmpcvndtlrltrgyern 456
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 15

AAE02782  
AAE02782 standard; Protein; 459 AA.

AC AAE02782;

DT 06-AUG-2001 (first entry)

DE Human six transmembrane epithelial antigen of prostate (STEAP)-3 protein.

DE Human: cytostatic; antiproliferative; vaccine; gene therapy;

KW six transmembrane epithelial antigen of the prostate-3; STEAP-3;

KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;

KW pancreatic.

OS Homo sapiens.

PN WO200140276-A2.

PD 07-JUN-2001.

PF 06-DEC-2000; 2000WO-US33040.

PR 06-DEC-1999; 99US-0455486.

PA (UROG-) UROGENESIS INC.

PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Paris M;

PI Jakobovits A;

DR WPI: 2001-367804/38.

DR N-PSDB: AAD07073.

PS Claim 36; Fig 10A-10B; 187pp; English.

The present sequence is human six transmembrane epithelial antigen of the prostate (STEAP)-3 protein. STEAP is a member of cell surface serpentine transmembrane antigens. STEAP-2 gene is located on chromosome 7q21 and is used in gene therapy. Inhibiting the development or progression of a cancer (eg, prostate, colon, bladder, lung, ovarian and pancreatic) expressing STEAP or inhibiting growth or killing cells expressing STEAP in a patient, comprises administering a vaccine composition to the patient. Treating a patient with a cancer that expresses STEAP, or inhibiting growth or killing cells expressing STEAP, comprises administering to the patient a vector encoding single chain monoclonal antibody that comprises the variable domains of the heavy and light chains of the monoclonal antibody that specifically binds to STEAP, such that the vector delivers the single chain monoclonal antibody coding sequence to the cancer cells and the encoded single chain monoclonal antibody is expressed intracellularly.

Note: The present sequence is stated as being the same as that shown as SEQ ID NO:10 (AAE02636) in sequence listing of the specification. However both the sequences differ at several positions.

```

Query Match 45.6%; Score 1155; DB 22; Length 459;
Best Local Similarity 46.6%; Pred. No. 2.3e-113;
Matches 214; Conservative 95; Mismatches 144; Indels 6; Gaps 3;

```



Result No.	Score	Query Match	Length	DB	ID	Description
1	160	6.3	198	2	D95285	conserved hypthet
2	159.5	6.3	211	2	AC2360	hypothetical prote
3	156.5	6.2	239	2	T50571	probable oxidoredu
4	147	5.8	213	2	H69400	conserved hypthet
5	142	5.6	212	2	D69361	conserved hypthet
6	142	5.6	224	2	T10120	F420-dependent NAD
7	133	5.2	223	2	D64487	hypothetical prote
8	121.5	4.8	629	2	S60385	probable membrane
9	121	4.8	191	2	P86626	hypothetical prote
10	120.5	4.8	232	2	A69131	conserved hypthet
11	119.5	4.7	216	2	T00121	hypothetical prote
12	116.5	4.6	242	2	G82642	conserved hypthet
13	116	4.6	198	2	AB3182	conserved hypthet
14	114.5	4.5	361	2	JC5653	G protein-coupled
15	112.5	4.4	365	2	S68208	G protein-coupled
16	112.5	4.4	1034	2	T30331	P-glycoprotein - T
17	111.5	4.4	1228	2	S59681	probable membrane
18	110.5	4.4	232	2	B84410	hypothetical prote
19	110	4.3	695	1	JN0898	foliitropin recept
20	109	4.3	320	2	T28379	ORF MSV218 hypthe
21	109	4.3	694	2	JC4301	foliitropin recept
22	109	4.3	1184	2	H71436	hypothetical prote
23	109	4.3	1301	2	D85188	disease resistance
24	108	4.3	712	2	S50969	probable membrane
25	107.5	4.2	202	2	B83059	hypothetical prote
26	107	4.2	1769	2	S53378	probable membrane
27	106.5	4.2	1442	2	B64582	sodium- and chlori
28	106.5	4.2	1695	2	JEO084	voltage-gated sodi
29	106	4.2	695	1	Q9H0F1	foliitropin recept

30	105	4.1	662	2	T31042	hypothetical prote
31	104.5	4.1	952	2	T05845	hypothetical prote
32	104	4.1	396	1	C64907	chloramphenicol re
33	104	4.1	396	2	G90895	hypothetical prote
34	104	4.1	396	2	H85721	probable resistanc
35	103.5	4.1	268	2	A70417	hypothetical prote
36	103.5	4.1	389	2	E96516	f16N3.13 [imported
37	103.5	4.1	465	2	S69915	sodium-phosphate t
38	103.5	4.1	521	2	T15322	hypothetical prote
39	102.5	4.0	314	2	H98310	probable peptide A
40	102.5	4.0	314	2	AB2972	hypothetical prote
41	102.5	4.0	714	2	AF2479	ABC transporter AT
42	102.5	4.0	1242	2	T39453	probable mrna stab
43	102	4.0	220	2	AG3547	bicyclomycin resis
44	101.5	4.0	1169	2	E97423	probable two-compo
45	101.5	4.0	1169	2	AC2641	hypothetical prote

## ALIGNMENTS

RESULT 1  
conserved hypothetical protein Sma0349 [imported] - Sinorhizobium meliloti (strain 1021)  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: D95285  
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; I  
C.: Kalmann, S.; Keating, D.H.; Palm, G.; Beck, M.C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A>Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: D95285  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <KUP>  
A:Cross-references: GB:AEO06469; PTDN:AKG64846.1; PID:g14523260; GSPDB:GN00165  
A:Experimental source: Strain 1021, megaplasmid pSymA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Publier, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
Pela, D.; Chailu, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalmann, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelanc  
nebaull, P.; Vandenberg, M.; Votholler, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen,  
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: Sma0349  
A:Genome: plasmid

Query Match	6.38; Score 160; DB 2; Length 198;
-------------	------------------------------------

Matches 55; Conservative 40; Mismatches 83; Indels 28; Gaps 8;

32 TVGVIGSGDEFAKSLTIRLCGYHVI--CSRNPKEAS---EFFPHVDVTHHEDALTKTN 87

db 3 TTAIGAGAIGSALAEFTAQIPAIANSRGPASSVTDRFGASVKAVELKDAL-qad 61

```
0Y      88 L I V A I H R E H Y T S L W D J R H L L V --- G K L L D V S N M R I N Q P ----- E S N A E Y L A S L F P   13
```

DD 02 VILLAV---PYDSIADI VIVQVSDMGGLVVDASNAIDEPAFKPRDEGGKSLIIVSELV IL

139 DELIVANV38NMWZLDFN D830KV11C31NN1ZNRWZV1ETNRWZNF1F1DLOSUS 139

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

```
Db      179 ASGPIQF-----GRPLVALNL 195
```

RESULT 2  
AC2560  
hypothetical protein alr8074 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AC2560  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchijima, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2560  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <KUR>  
A:Cross-references: GB:AP003603; PIDN:BA877404.1; PID:gl7134847; GSPDB:GN00182  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr8074  
A:Genome: plasmid

Query Match 6.3%; Score 159.5; DB 2; Length 211;  
Best Local Similarity 24.9%; Pred. No. 3.1e-05;  
Matches 51; Conservative 43; Mismatches 70; Indels 41; Gaps 8;  
QY 33 VGVIGSGDFAKSLIRLCRGVHVIGSRNP-----KFASEFPFHVDVTHHEDALTKTN 87  
DB 3 IAFIGIGVGSALASQLSLDHTVITARNSDSVKALKKYPE-LOVSSPQEAIAQAE 61  
QY 88 IIFVA---IHREHYTSLMDRLHLVGLKILIDVSNMNRIN---QYPSNAEYLASLFP 138  
DB 62 VIFLPTPTTAQALAEVGDLS---GKILVDTCTPVGNTLHGKSPQSEILOSVP 117  
QY 139 DSLIVKGFNVYSANALQGPDKASROY-----ICSNNOAQOYIELARQINF 187  
DB 118 HAKVAKATTI-----YGFNFENHTYPGYGNLKPAMLIAGNDVPKQVSTLCQQLGW 170  
QY 188 IPIDGLSSAREIENLPRLFTLM 212  
DB 171 EAVDVGNLSMHLHEHMTL-----LM 191

RESULT 3  
T50571  
probable oxidoreductase [imported] - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 28-Jul-2000  
C:Accession: T50571  
R:Redenbach, M.; Kleiser, H.M.; Denapalte, D.; Eichner, A.; Cullum, J.; Kinashl, H.; Hopa Mol. Microbiol. 21, 77-96, 1996  
A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S. coelicolor chromosome  
A:Reference number: Z20556; MUID:97000351  
A:Accession: T50571  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-239 <RED>  
A:Cross-references: EMBL:AL133220; PIDN:CAB61708.1  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Note: SCC75A.08c  
C:Superfamily: conserved hypothetical protein MJ1501

Query Match 6.2%; Score 156.5; DB 2; Length 239;  
Best Local Similarity 27.9%; Pred. No. 6.2e-05;  
Matches 61; Conservative 41; Mismatches 72; Indels 45; Gaps 10;  
QY 25 IKDKRKYTVGVI-GSGGFAKSLIRLCRGVHVIGSR---NPKFASEFPFHVDVTHHE 80  
DB 23 LPDVSGLVGVGTFQGGKGLATLAKAGKATVGSAAERAAAEIIGHVEGADNA 82

QY 81 DALTKNTIIFVAIHRE-HYTSIMDLRLHLVGLKILIDVSNMNRINO-----YPE--SNAE 131  
DB 83 ETARSDVVIYAVPWGDKLTLESIRALSGKIVDVCNPLGFDKCAVALKEEGSAAE 142  
QY 132 YLASFPDLSLVKGFNVYSANALQ-----LGPDKASROYICSNNOAQOYIELARQINF 178  
DB 143 OAAALLPDSRAAAAFHLSAVLLQDPEIDETDVMVAGEERADVEI-----VOA----- 192  
QY 179 IEIARQINFIP---IDLGSLSAREIENLPRLFTLM 213  
DB 193 --LAGR---IFGMKGVFAGRLRNHMQVESIVANLISVNR 226

RESULT 4  
H69400  
conserved hypothetical protein AF1209 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: H69400  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing, arch.  
A:Reference number: A69250; MUID:98049343  
A:Accession: H69400  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-213 <KLE>  
A:Cross-references: GB:AE01021; GB:AE00782; NID:92669344; PIDN:AAE90038.1; PID:9264  
C:Superfamily: conserved hypothetical protein MJ1501

Query Match 5.8%; Score 147; DB 2; Length 213;  
Best Local Similarity 27.5%; Pred. No. 0.0003;  
Matches 56; Conservative 44; Mismatches 68; Indels 36; Gaps 11;  
QY 37 GSGDFAKSLIRLCRGVHVIGSRNP-----KFASEFPFHVDVTHHEDALTKTN 87  
DB 8 GTGNIGEGELARWGLKLTGIVTSRKLKAEKASLDYKAGDASIIIGKREDAAETCD 66  
QY 89 IFVAIHREHYTSLMD-----LRHLVGLKILID-----VSNMNRINOYPE--SNAE 131  
DB 67 --VAV---FTIPWFAFDPAEMLRQLAGKIVSLPVMKQVDNF--VYVRPEGSAAE 119  
QY 132 YLASFPDLSLVKGFNVYSANALQGPDKASROYICSNNOAQOYIELARQINF 190  
DB 120 KLASVLESSVVAAYHSLIPARRFANLGEFEMDVPICGS-GAREVYVDLTKISGLRAL 178  
QY 191 DLGSLSSAREIENLPRLFTLM 213  
DB 179 DAGLSNAHLVESLTPILINMKR 202

RESULT 5  
D69361  
conserved hypothetical protein AF0892 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: D69361  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch.  
A:Reference number: A69250; MUID:98049343  
A:Accession: D69361  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA







DB 126 AAKANVVKCFSSVGS-GLAVNPOLKGEKPSMFICGNDSSKQIKIIDLTFMGDMEDMK 184  
OY 195 LSSAREIENPLRLFTLW 212  
DB 185 VEARAIEPLCI---LW 198

## RESULT 12

conserved hypothetical protein XF1737 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: G82642

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
X:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
X:Reference number: AB2515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: G82642

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-242 <SIM>  
A:Cross-references: GB:AE003997; GB:AE003849; NID:99106805; PIDN:AAF4546.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
B:Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.F.; Ferreira, A.J.S.  
submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kempfer, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H  
A:Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sancelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1737

Query Match 4.6%; Score 116.5; DB 2; Length 242;

Best Local Similarity 24.0%; Pred. No. 0.089;  
Matches 46; Conservative 40; Mismatches 85; Indels 21; Gaps 7;

OY 28 ARKVTGIGSGDFAKSLTIRLCGYHVVIGSRNP---KFASEFPFHYVDVTHHEDAL 83  
DB 30 AAPMRIGVIGAGSLGTGVLWYKAGHEVWFSRNPDKLEAMARELEPR-ASVGQPLAAT 88  
OY 84 TKTNIIFVAIHREHYSLM-DLRHLVYKILIDVSNMNMINO---YPSRN---AEYLAS 135  
DB 89 EFGYVLLAVPEALPQVGRDLSAYRKIVLDTNPMGCASSADVYREARELGVAACTVVK 148  
OY 136 LPDPSLIVGFNVASAWALOLGPKDASRO---VYICSNNTIOAROVIELARQLFTPI 190  
DB 149 YMPGALVAFSAVDATVY---TSASRGRGRLGMLASDDAEAMKVAAGLVADACDPV 205  
OY 191 DLGSLSSAREIE 202  
DB 206 IVCNLAAMASFQ 217

## RESULT 13

conserved hypothetical protein Atu5183 [imported] - Agrobacterium tumefaciens (strain C5  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AB3182

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kan  
ster, E.W.  
X:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
X:Reference number: AB2577; PMID:11743193  
A:Accession: AB3182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <KUR>

A:Cross-references: GB:AE008687; PIDN:AA45872.1; PID:q17743615; GSPDB:GN00188  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu5183  
A:Genome: plasmid

Query Match 4.6%; Score 116; DB 2; Length 198;  
Best Local Similarity 22.7%; Pred. No. 0.076;  
Matches 50; Conservative 41; Mismatches 81; Indels 48; Gaps 10;

OY 31 VYVGVIGSGDFAKSLTIRLCGYHYI-GSRPKRASEFFPV---VDVTHHEDALTKT 86  
DB 1 MTVGIIGAGNIGAFATGACIGAEVIANSRGPELSLALVSKLSTTRAGSVPEAAQA 60  
OY 87 NIFVAIHREHYSLMD---LRHLVG-KILIDVSNMNMINOY-----PESNAEYLA 134  
DB 61 -YLVAVP-----WSKIPGALAGLFGDRIVYIDANNISLEAPLYRPADLGRTSDIFF 112  
OY 135 LFPDPSLIVGFNVASAWALOLGPKDAS-----ROYICSNNTIOAROVIELARQL 185  
DB 113 ALTPGARVYKAFN-----HLTPKOLSGDPHSEGRVLYFGSDMMRAKAEVGAIDRI 165  
OY 186 NFPIIDLSLSSAREIENPLRLFTLRGPVYVAISLAF 225  
DB 166 GFGIDLGGLPGVSQLLQFP-----GSPPALNLVKT 197

## RESULT 14

JC5653

G protein-coupled receptor - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 17-Nov-1997 #sequence\_revision 17-Nov-1997 #text\_change 21-Jul-2000  
C:Accession: JC5653

R:Ferrer-Martinez, A.; Felipe, A.; Mata, J.F.; Casado, F.J.; Pastor-Anglada, M.  
Biochem. Biophys. Res. Commun. 238, 107-112, 1997  
X:Title: Molecular cloning of a bovine renal G-protein coupled receptor gene (bGR):  
A:Reference number: JC5653; MUID:97445134  
A:Accession: JC5653  
A:Molecule type: mRNA  
A:Residues: 1-361 <FER>

A:Cross-references: GB:U88366; NID:92827875; PIDN:AAC05611.1; PID:92827876  
A:Experimental source: renal epithelial cell  
C:Superfamily: G protein-coupled receptor 4  
C:Keywords: glycoprotein; phosphoprotein  
F:24-46/Domain: transmembrane #status predicted <TM1>  
F:58-79/Domain: transmembrane #status predicted <TM2>  
F:96-117/Domain: transmembrane #status predicted <TM3>  
F:137-159/Domain: transmembrane #status predicted <TM4>  
F:189-207/Domain: transmembrane #status predicted <TM5>  
F:229-253/Domain: transmembrane #status predicted <TM6>  
F:265-289/Domain: transmembrane #status predicted <TM7>  
F:318-62/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:134,218,293,313/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status  
F:204,221,328/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status

Query Match 4.5%; Score 114.5; DB 2; Length 361;  
Best Local Similarity 20.4%; Pred. No. 0.21;  
Matches 56; Conservative 51; Mismatches 86; Indels 81; Gaps 13;

OY 255 VNKTL-PIVATITLSLVYLAGLIAAAYQLYGTYKRRFPPELTWQCRKQGLSFFFA 313  
DB 19 IHQTLAPVYV---WLVAVGFPAACLSTLYG-----YLD--IKARRELGVYLCNLT 64

```

OY 314 MVAHVASLTCLPMMRBSERLFLNMAVQOVHANIEANSNEE-----EVMNIEKYSIFGMSIG 369
Db 65 VADFLYICSLP-----FWLQYVLQHDH-----WSDHDSOCVCGILYENIYI-SVG 110
OY 370 LLSLLAVTSPISVSNALMRBSFSPISQSLTGVAL-----LISFPHVLIY-----GWK 416
Db 111 FLCCSIDIRYLAVAHPRFQHGFRLTKAMGSVALIWKELTTSYFLMBEVEDADRHR 170
OY 417 RAFE-----EYRRFTPPNFVALVLSI-----VI 443
Db 171 VCFEHYPLEPRGRGINVYRFLVGLFPLICLLLASRGLRAVRRSHOTOKSRKDOJORLV 230
OY 444 LGGIILFLPCL--SRKLKRIKKGGKESQPLDEGI 475
Db 231 LSTVIVIFLACFLPYHVLILVNSLMESSCDRAKI 264

```

Gencore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2002, 02:12:50 ; Search time 37.74 Seconds

(without alignments)  
502.718 Million cell updates/sec

Title: US-09-802-520-1

Perfect score: 2534

Sequence: 1 MESISMWGSKSLSETCLPN.....LEEGIGTIPHSPEPTYM 490

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	31.1	339	1 STEA_HUMAN	Q58896 homo sapien
2	133	5.2	223	2 FARE_METJA	058896 methanococc
3	121.5	4.8	629	1 FARE_YEAST	012333 saccharomyc
4	120.5	4.8	232	1 FARE_METTH	026350 methanobact
5	118.5	4.7	365	1 GPE8_HUMAN	015743 homo sapien
6	110	4.3	695	1 FSHR_MACFA	P32212 macaca fasc
7	108	4.3	712	1 FPRE6_YEAST	Q12473 saccharomyc
8	107	4.2	1769	1 YUK9_YEAST	P42945 saccharomyc
9	106	4.2	695	1 FSHR_HUMAN	P23945 homo sapien
10	104	4.1	396	1 SOTB_ECO57	P56529 escherichia
11	104	4.1	396	1 SOTB_ECOLI	P31122 escherichia
12	103.5	4.1	268	1 YD49_AQUAE	067364 aquifex aeo
13	103.5	4.1	465	1 NPPI_MOUSE	Q61983 mus musculu
14	103.5	4.1	521	1 YR25_CAEEL	Q10934 caenorhabdi
15	103	4.1	695	1 FSHR_PIG	P45059 sus scrofa
16	102	4.0	693	1 FSHR_CHICK	P79763 gallus gali
17	101	4.0	393	1 NOCC_ARATH	P56753 arabidopsis
18	101	4.0	393	1 NOCC_SPIOL	Q56315 spinacia ol
19	100	3.9	476	1 RF3_SACBA	P77409 saccharomyc
20	99	3.9	458	1 SSUL_YEAST	P41930 saccharomyc
21	98.5	3.9	452	1 NUD4M_BRALA	079421 branchiosto
22	98.5	3.9	3411	1 POLG_YEYV1	P03314 y genome po
23	98.5	3.9	3411	1 POLG_YEYV2	P19901 y genome po
24	96.5	3.8	686	1 FREL_YEAST	P33791 saccharomyc
25	96	3.8	261	1 PHSC_ECOLI	P77409 saccharomyc
26	96	3.8	447	1 NUD4M_APILI	P34853 apis mellif
27	96	3.8	695	1 FSHR_BOVIN	P33376 bos taurus
28	95.5	3.8	393	1 NUDC_ORYSA	P12132 oryza sativ
29	94.5	3.7	433	1 NUD4M_RHISA	098825 rhizopcephal
30	94.5	3.7	452	1 NUD4M_BRALF	047423 branchiosto
31	94.5	3.7	492	1 SECY_CYAPA	P25014 cyanophora
32	94.5	3.7	1228	1 ATC_PLAFK	008853 plasmodium
33	94	3.7	891	1 VP4A_FOWPV	Q93559 fowlpox vir

34	93.5	3.7	499	1 GSHR_PLAF7	015770 plasmodium
35	93.5	3.7	1233	1 VIL1_REOVD	P15024 reovirus (t
36	93	3.7	388	1 HMC3_DESVH	P33390 desulfovibr
37	93	3.7	564	1 HXTC_YEAST	P39924 saccharomyc
38	93	3.7	692	1 FSHR_RAT	P20395 rattus norv
39	93	3.7	695	1 FSHR_SHEEP	P35379 ovis aries
40	92.5	3.7	379	1 CYB_MIRLE	Q35019 mirounga le
41	92.5	3.7	416	1 YH04_AQUAE	067601 aquifex aeo
42	92.5	3.7	477	1 BAIG_EUBSP	P32369 eubacterium
43	92.5	3.7	490	1 GTRL_CHICK	P46896 gallus gall
44	92.5	3.7	1120	1 KEPA_ECOLI	P77338 escherichia
45	92.5	3.7	2037	1 FAS1_CANAL	P34731 c fatty aci

## ALIGNMENTS

RESULT ID	STEA_HUMAN	STANDARD:	PRT:	339 AA.
AC	Q9UEH8: Q95034:			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Six transmembrane epithelial antigen of prostate.			
GN	STRAEP OR STRAEP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid:9606;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE:20056277; PubMed-10588738;			
RA	Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,			
RA	Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,			
RT	Jakovovits A., Saffran D.C., Afar D.E.H.;			
RT	"STRAEP: a prostate-specific cell-surface antigen highly expressed in			
RT	human prostate tumors";			
RT	Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Abu-Threideh J., Stoneking T., Langston Y., Maupin R.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL: AF186249; AAF17479.1; -			
DR	EMBL: AC005053; AAC79150.1; ALT_INT.			
DR	EMBL: AC004969; AAD15620.1; ALT_INT.			
DR	MIM: 604415; -			
KW	Transmembrane; Antigen.			
FT	TRANSMEM 71 91	POTENTIAL.		
FT	TRANSMEM 119 139	POTENTIAL.		
FT	TRANSMEM 164 184	POTENTIAL.		
FT	TRANSMEM 218 238	POTENTIAL.		
FT	TRANSMEM 258 278	POTENTIAL.		
FT	TRANSMEM 291 311	POTENTIAL.		
FT	SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;			
QY	208	LEFTINGPVPVVAISLSTFFFLVSFVADVHPVARNQSDPFYKIPIDLVKTLPIVAITLL 267		

Query Match 31.1%; Score 788; DB 1; Length 339;  
Best Local Similarity 54.8%; Pred. No. 4,9e-48;  
Matches 142; Conservative 53; Mismatches 64; Indels 0; Gaps 0;

Db 67 LFPQMHPIKIAIATIASTFLTYLREVIHPLASTSHOQYFKIKILVINKLPMVSTLL 126  
 Oy 268 SLVYLAGLAAAYOLYXGTGKRRPPEMTLOQRKQGLISFFPAHVAVSLCLPARR 327  
 Db 127 ALVLPQVIAIYVOLHNGTKYKRRPPEMTLOQRKQGLISFFPAHVAVSLCLPARR 186  
 Oy 328 SERFLPMAYQVYHANIENSNNEEYWRLEMYISFGIMSLGLSLAVTSPISVSNALN 387  
 Db 187 SYRKLLNMAVYQVQOKKEDMIEDHWMEIYVSLGVIALALAVTSIPSVDSLT 246  
 Oy 388 WRESFQSLGIVALLISFPHVILYKRAFEERYRFTPPNFVALVLPSTVILGKI 447  
 Db 247 WREHHYQSKIGIVSLIGTILHALIFAMNKWIDIKOFVWPTPTMIAVFLPVLIFKS 306  
 Oy 448 ILFLPCISRKLRKIKGME 466  
 Db 307 ILFLPCIRKIKLRKIGME 325

## RESULT 2

F4RE.METJA STANDARD; PRT; 223 AA.  
 ID F4RE.METJA STANDARD; PRT; 223 AA.  
 AC Q58886;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative F420-dependent NADP reductase (EC 1.-.-.-).  
 GN M1501.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=9637999; PubMed=8688087;  
 RA Sait C.G., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sult G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kierleaker A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhman J.L., Nguyen D.,  
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotten M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RT Science 273:1058-1073(1996).  
 RL Science 273:1058-1073(1996).  
 CC - FUNCTION: CATALYZES THE REDUCTION OF F420 WITH NADP(+) AND THE  
 CC REDUCTION OF NADP(+) WITH F420H(2). PROBABLY COUPLES THE NADP-  
 CC DEPENDENT OXIDATION OF THE ALCOHOL TO THE ALDEHYDE WITH DE F420-  
 CC DEPENDENT REDUCTION OF CO(2) TO METHANE (ANABOLIC FUNCTION) (BY  
 CC SIMILARITY).  
 CC - CATALYTIC ACTIVITY: NADPH + COENZYME F420 = NADP(+) + REDUCED  
 CC COENZYME F420.  
 CC - SIMILARITY: TO M.ORGANOPHILUM F420-DEPENDENT NADP REDUCTASE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U67591; AAB99514.1; -;  
 DR TIGR; M1501; -;  
 KW Hypothetical protein; Oxidoreductase; NADP; Complete proteome.  
 SO SEQUENCE 223 AA; 24068 MW; 2370BBD6F5BBD3D9 CRC64;

Query Match 5.2%; Score 133; DB 1; Length 223;  
 Best local similarity 21.9%; Pred. No. 0.013;

Matches 46; Conservative 52; Mismatches 78; Indels 34; Gaps 7;  
 Oy 37 GSGDFAKSLFTRLRGCGHVHVGSRNPFASFEPHYVDV-----THHEDLT 84  
 Db 8 GTDQGGELAIRLAK-NKKITIGSRKKKAFAEAKKAKKEILKONGIENDITGLEKNDAAK 66  
 Oy 85 KTNIIFAVHREH-YTSLMDLRHLVGLKILDV-----SNMRINOPE-SNAEYL 133  
 Db 67 EGDVYILSLPEYETLSTIKQLKEELKGIYVIGVPLATAGDKPTRLFPDGSVAEMV 126  
 Oy 134 ASLPDLSLVGVGVNVSMAALQLOPKDASROYVICSNNIARQOVIELARLNTI-PIDL 192  
 Db 127 QNVAKESKVSFAFQNVCHAVIEDIDNPVDCDILVCGNDEAKKVIDLANIDGVRAIDC 186  
 Oy 193 GSLSSAREENLPLRLFTLMRGVVAISL 222  
 Db 187 GNLEKSRILEAIT-----PLIGLINT 207

## RESULT 3

FRE7\_YEAST STANDARD; PRT; 629 AA.  
 ID FRE7\_YEAST STANDARD; PRT; 629 AA.  
 AC Q12333;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ferric reductase transmembrane component 7 (EC 1.6.99.13) (Ferric-  
 DE chelate reductase 7).  
 GN FRE7 OR YOL152W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=96132030; PubMed=8553699;  
 RA Casamayor A., Aldea M., Casas C., Herrero E., Gamo F.J.,  
 RA Lafuente M.J., Gancedo C., Arino J.;  
 RT "DNA sequence analysis of a 13 kbp fragment of the left arm of yeast  
 RT chromosome XV containing seven new open reading frames";  
 RL Yeast 11:1281-1286(1995).  
 CC - CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).  
 CC - COFACTOR: FAD (PROBABLE).  
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC - SIMILARITY: BELONGS TO THE FRE / CYB8 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z48239; CAAB8276.1; -;  
 DR EMBL; Z74894; CAA99174.1; -;  
 DR SGD; S0005512; FRE7.  
 DR InterPro: IPR002916; Ferric\_reduct.  
 DR Pfam: PF01794; Ferric\_reduct.1.  
 KW Oxidoreductase; Electron transport; Transmembrane; Iron transport;  
 KW FAD; NAD; Glycoprotein; Multigene family.  
 FT NP\_BIND 369 375 FAD (POTENTIAL).  
 FT TRANSMEM 46 66 POTENTIAL.  
 FT TRANSMEM 108 128 POTENTIAL.  
 FT TRANSMEM 168 188 POTENTIAL.  
 FT TRANSMEM 195 215 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 266 286 POTENTIAL.  
 FT TRANSMEM 293 313 POTENTIAL.  
 FT TRANSMEM 422 442 POTENTIAL.  
 FT TRANSMEM 442 462 POTENTIAL.  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 541 541 N-LINKED (GLCNAC. .) (POTENTIAL).

**SEQUENCE** 629 AA; 71996 MW; 2384480E289C16P CNC64;

Query Match 4.8%, Score 121.5; DB 1; Length 629;  
Best Local Similarity 21.1%; Pred. No. 0.28;  
Matches 63; Conservative 55; Mismatches 88; Indels 93; Gaps 19;

OY 214 GPVVAIAISLAFPFILYSVRDVIHPHYARNOQSDFKPIPELYNKTPLPIVAITLLSLV-L 272  
| : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 123 GFELVMATTTLYTLKCEVP---HFYYR-PCAGFSSPLSY--RAGIMAISSLVPEFSL 175  
OY 273 AG-----LLAAAYQLYGTGYRRFPWLEMTLQCRQLGLSPFFMWHAVYSLCPM 325  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 176 SGKINVCIMVLGSLTE-----KINITYHOM-----ASTLCFEFSWHA-----IPE 215  
OY 326 RRSRETYFLANNAQOVHANINENSNEEEWRHEMYTISFGISLGLSILAAYSIP----- 380  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 216 LRQARRH---EGGEYEMH----QRMKASDMNR-----SGVFPLEFLNLMLSSLP IARRH 262  
OY 381 -SVSNALNMREFESTIGTGLVALLISTFHV-----LIYMKRAFEDEYYR- 425  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 263 FYEFLQLQHW-----ILAVGFYISLF--YHYPELNSHMVLAIVTW---FAOLFVRL 311  
OY 426 ----FYTPPNRYLVALVEPSIVYLGGIIIFLPDISIKKLRIKKGWKSQP-----LEEGI 475  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 312 AVKGALRGSRFSFMASTIANVSIVE-----GCVELIVDVEMAYSPGOHIIFVRTDKGI 365

**RESULT** 4  
**FARE.METTH** STANDARD; PRT; 232 AA.  
AC O26350;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative F420-dependent NADP reductase (EC 1.-.-.-).  
GN MTH248.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OX Methanothermobacter.  
NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
Harrison D., Hoang L., Keagle P., Lum W., Potlter B., Qiu D.,  
Spadatore R., Viciare R., Wang Y., Wierzbowski J., Gibson R.,  
Jivani N.N., Carruso A., Bush D., Saffer H., Patwell D., Prabhakar S.,  
McDonough S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
Daniels C.J., Mo J.-I., Rice P., Neelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltah: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -1- FUNCTION: CATALYZES THE REDUCTION OF F420 WITH NADP(+) AND THE  
REDUCTION OF NADP(+) WITH F420H(2). PROBABLY COUPLES THE NADP-  
DEPENDENT OXIDATION OF THE ALCOHOL TO THE ALDEHYDE WITH DE F420-  
DEPENDENT REDUCTION OF CO(2) TO METHANE (ANABOLIC FUNCTION) (BY  
SIMILIARITY).  
CC -1- CATALYTIC ACTIVITY: NADPH + COENZYME F420 = NADP(+) + REDUCED  
COENZYME F420.  
CC -1- SIMILARITY: TO M.ORGANOPHILUM F420-DEPENDENT NADP REDUCTASE.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

OR EMBL; AE000811; AAB84754.1; -

```

KW Hypothetical protein: Oxidoreductase, NMDP; Complete proteome.
SQ SEQUENCE 232 AA; 24539 MW; AICE60ABC8474296 CRC64;

Query Match 4.8%; Score 120.5; DB 1; Length 232;
Best Local Similarity 23.8%; Pred. No. 0.1;
Matches 49; Conservative 42; Mismatches 90; Indels 25; Gaps 6.

OY 30 KVTGVIGSGDFAKSLTIRLRCGYHYHVGISFNPFASEFPYVDVTHHE----- 80
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 8 KIAV-IGGTGDDGLALRFVAAGEEVILIGSDAKASKASKAEIAGRDISVEGATN 66
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 81 -DALTKNIIIPAIH-REHYTSLMDLRHLVAKILLDYS-----NNMRINQYPSRN 129
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 67 PDAAASADVVYVLTPLQAMWTPLASIRQYRDKVILIDATVPIDSCIGSAVRIDMEGS 126
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 130 AEYLASLF---PDSLTIVGFENVVSAMALDLPKDSRQYICSNNTIOARQVYTELARQIN 186
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 AAERARPLRREGCTRYAAAFNNISASALLEVSEPDCCCLVASDRHDALEVAELAELAKID 186
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 187 FI-PIDGLSSLSAREIENLPRLFTL 211
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 GYRAIEGGLGNARLIEKTPILLINL 212
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 5
GP68_HUMAN STANDARD; PRT; 365 AA.
ID GP68 HUMAN
AC Q15743; O13334;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable G protein-coupled receptor GPR68 (Ovarian cancer G protein-
   coupled receptor 1) (OCR-1).
GN GPR68 OR OGR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA MEDLINE=96299795; PubMed=8661159;
RX Xu Y., Casey G.;
RT "Identification of human OGR1, a novel G protein-coupled receptor
   that maps to chromosome 14.";
RL Genomics 35:397-402(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA An S., Goetzl E.J.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: ORPHAN RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: FOUND IN MANY TISSUES, INCLUDING SPLEEN,
   TESTIS, PERIPHERAL BLOOD LEUCOCYTES, BRAIN, LUNG, AND PLACENTA.
CC NOT DETECTABLE IN OVARY.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U48405; AAC50596.1; -.
DR EMBL; U35398; AAA79060.1; -.
DR GCRDB; GCR_1937; -.
DR GCRDB; GCR_1960; -.
DR MIM; 601404; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.

```

DR PRINTS: PR00237; GPCR\_Rhodopsin.  
 DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 DR PROSITE; PS50262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 21  
 FT TRANSSEM 22 46  
 FT DOMAIN 47 58  
 FT TRANSSEM 59 80  
 FT DOMAIN 81 95  
 FT TRANSSEM 96 117  
 FT DOMAIN 118 136  
 FT TRANSSEM 137 158  
 FT DOMAIN 159 183  
 FT TRANSSEM 184 203  
 FT DOMAIN 206 228  
 FT TRANSSEM 229 249  
 FT DOMAIN 250 263  
 FT TRANSSEM 264 284  
 FT DOMAIN 285 365  
 FT DISULFID 94 172  
 FT CARBOHYD 3 3  
 FT CARBOHYD 8 8  
 FT CONFLICT 140 140  
 FT CONFLICT 142 142  
 FT CONFLICT 142 142  
 SQ SEQUENCE 365 AA; 41076 MW; 05919AFD5B842CCD CRC64;

Query Match 4.7%; Score 118.5; DB 1; Length 365;  
 Best Local Similarity 21.0%; Pred. No. 0.24;  
 Matches 57; Conservative 46; Mismatches 92; Indels 77; Gaps 11;

QY 255 VNKL-FIVATITLSLYAGLAATQVLYGTYRRPPLEWLOCRKGLISFFFA 313  
 DB 19 IHOGLAVVYVTV-----VVGFPANCLSLYFG-----YLQ--IKARNELGYLNL 64  
 QY 314 MVHAYSLCLPMRSEKRLFLNMAVQOVHANIEENWNEEVRMIEISFGIMSLISTL 373  
 DB 65 VADLFYICSLPFW-----LQYVLOHDMWHSGLSCQVGLILENYIY-SVGFLLCC 114  
 QY 374 LAVYSIPSVNALNMRSEFISITLGYVA-----LLISTP----- 408  
 DB 115 ISVDRYLAHVHPRFHQFRIKAAGVSVMWAKELLSITFLMHEVIEDENGRVCFE 174  
 QY 409 HVLVYGMKRAFEERYRPPNFVLAIVPST-----VILG 445  
 DB 175 HYPLOAMQRAI--NYRFLVGLFPLICLLASVYGLIARVSHGTOKSRKQIQRVLIS 232  
 QY 446 KIILFPCI--SRKLRIKGMKESQFLBEGI 475  
 DB 233 TVVIFLACPLPVHLLVRSVWASCDPAKGV 264

RESULT 6  
 FSHR\_MACFA STANDARD; PRT; 695 AA.  
 AC P32212;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE FSHR follicle stimulating hormone receptor precursor (FSH-R) (Follicotropin receptor).  
 GN FSHR.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=94071854; PubMed=7504463;  
 RA Gromoll J., Dankbar B., Sharma R.S., Nieschlag E.;  
 "Molecular cloning of the testicular follicle stimulating hormone

RT receptor of the non human primate Macaca fascicularis and  
 RT Identification of multiple transcripts in the testis.;  
 RL Biochem. Biophys. Res. Commun. 196;1066-1072(1993).  
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC FSH/LSH/TSR SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X74454; CA52463.1; -.  
 CC PIR; S36452; S36452.  
 CC PIR; JN0898; JN0898.  
 CC HSSP; P23945; 1XUN.  
 CC GCRD; GCR\_0653; -.  
 DR InterPro; IPR000276; GPCR\_Rhodopsin.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000372; LRR\_Nterm.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR Pfam; PF00560; LRR; 4.  
 DR Pfam; PF01462; LRRNT; 1.  
 DR PRINTS; PR00373; GLXCHROMER.  
 DR PRINTS; PR01143; FSHRECEPTOR.  
 DR SMART; SM00013; LRRNT; 1.  
 DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 DR PROSITE; PS50262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; signal;  
 KW phosphorylation; Repeat; Leucine-rich repeat.  
 FT SIGNAL 1 17  
 FT CHAIN 18 695  
 FT DOMAIN 18 366  
 FT TRANSSEM 367 387  
 FT DOMAIN 388 398  
 FT TRANSSEM 399 421  
 FT DOMAIN 422 443  
 FT TRANSSEM 444 465  
 FT DOMAIN 466 485  
 FT TRANSSEM 486 508  
 FT DOMAIN 509 529  
 FT TRANSSEM 530 550  
 FT DOMAIN 551 573  
 FT TRANSSEM 574 597  
 FT DOMAIN 598 608  
 FT TRANSSEM 609 630  
 FT DOMAIN 631 695  
 FT REPEAT 44 68  
 FT REPEAT 69 93  
 FT REPEAT 119 143  
 FT REPEAT 170 192  
 FT REPEAT 193 216  
 FT REPEAT 218 240  
 FT DISULFID 442 517  
 FT CARBOHYD 191 191  
 FT CARBOHYD 199 199  
 FT CARBOHYD 293 293  
 FT CARBOHYD 318 318  
 SQ SEQUENCE 695 AA; 78343 MW; 0D60A233729B5250 CRC64;

Query Match 4.3%; Score 110; DB 1; Length 695;  
 Best Local Similarity 16.5%; Pred. No. 2;  
 Matches 84; Conservative 79; Mismatches 151; Indels 196; Gaps 17;

QY 79 HEDALTKTNIIFVAIHREHYTSLMDLRHLLVG-----KILIDVSNMNR 121



```

Db 98 HEIRIKKANL-LYINPEAFONLPNRLYLISNTGIKHLPDVAKHISFOKVLIDIDNNIN 156
OY 122 INQPESSNAEYLASFPPDSLIVKGFENVVSAMALQLOPKDASROYICSNNTICARQOYIEL 181
Db 157 IH-----TIERNSEFVGLSFESVILMT-----NKGIOEIHNC----- 189
OY 182 ARQLNFPIDIGLSSAREIENLPLRLFTLMRGPPVVAISLAEFFLYSPVRDIHPYAR 241
Db 190 ----FNCTQDELNLSDNNNLDELPLNDVFHGASGPVILDISIRIHSPLSIGLENKLKLA 246
OY 242 NOOSDFYKIPRIEIVNKTPIVAITLLSLVY----- 271
Db 247 RSTYNNLKLP-----SLEKVALMEASLYFPSHCAFAMWRQISLHPICNKSILROEV 301
OY 272 -----LAGLLAAAYQLYGTGKRRF-----PPMLLETWLOCKROIG 306
Db 302 DVMOTRGORSSLAEDNESSYSRNGFDMTAEFDYDLCEVVDVTCSPKPDAPNCPEDIIIG 361
OY 307 -----LSFPPAMVHYA-----YSICLPMRSEERYFLNMAVYQOVHANIENS 348
Db 362 YNLRVLIWFISLITATGNIIVLVLTTSQYKLTVP-----RFLMCNLAFAD----- 408
OY 349 MNEEYWRITEMYISFGIMSLGLSLAVTSIPSVSNALNMR-----FSFIQSTIG 399
Db 409 -----LCIGIYLLLASVDIHTKSOYHNYAIDMQGAGCDAGFTVFASELS 456
OY 400 ----YVALLISTFH-----VLIYGRAPREEEYRYRTPNEY-LAL 436
Db 457 VYTLAATLELRWHTTHAMQLOCKVHRHAAVSVMGWTFAPFAALFPIFGISSYMKVSI 516
OY 437 VLP-----SIVILGKIILFLPC 453
Db 517 CLPMDISPLSQLYMSLIVLVNLAFAVIC 546

```

```

RESULT 7
FREQ_YEAST STANDARD: PRT: 712 AA.
AC 012473:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric reductase transmembrane component 6 precursor (EC 1.6.99.13)
GN FERRIC-CHELATE REDUCTASE 6.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RA STRAIN=S288C;
RC Medler H., Wandutt R.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC -1 COFACTOR: FAD (PROBABLE).
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1 SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 247973; CAA88006.1; -
DR EMBL: 273156; CAA97503.1; -
DR SGD: S0003974; FRE6.
DR InterPro: IPR002916; Ferric_reduct.
DR Pfam: PF01794; Ferric_reduct; 1.

```

```

KM Oxidoreductase; Electron transport; Transmembrane; Iron transport;
KW FAD; NAD; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 17
FT CHAIN 18 712
FT NP_BIND 493 499 FAD (POTENTIAL).
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 478 498 POTENTIAL.
FT TRANSMEM 550 570 POTENTIAL.
FT DOMAIN 553 576 POLY-LEU.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 712 AA; 81969 MW; 5224F12B51544BAA CRC64;

```

Query Match 4.3%; Score 108; DB 1; Length 712;  
 Best Local Similarity 18.5%; Pred. No. 2.8;  
 Matches 74; Conservative 52; Mismatches 114; Indels 160; Gaps 17;

```

OY 88 IIFVAIHREHYTSL-----WDLRLVLGKILIDVSNMNRINOYPESNAEYLAL 136
Db 183 IIAVFPMHSHYNGLNALRAFASREFVNYIRGHVLPTELYD----KHANHEKFLNVEFTGL 238
OY 137 FPDLSLVKGFENVVSAMALQLOPKDASROYICSNNTICARQOYIELARQLNFPIDIGSL 196
Db 239 MPNSL-----EAW----- 246
OY 197 SAREIENLPLRLFTLMRGPPVVAISLAEFFEL-YSEVRDIHPYARNQOSDFYKIPRIEIV 255
Db 247 -----IIRGYTLANIIFLISISIIIDEPYNNLIFNSHLSQFRL---LA 284
OY 256 NKTPIVAITLLSLVYLAGLLAAAYQLYGTGKRRPPLMETWLOCKROIGLISFPPAMY 315
Db 285 DRS-GILAFLOPFLIIIFARNSFLEFLGVKKNF-----ISPHKMGIRIMVNAFI 336
OY 316 H-VAYSICLPMRSEERYFLNMAVYQOVHANIENSNAEYWRITEMYISGIMSLGLSL 374
Db 337 HSLSYSL-----FAIINNAFK-----ISNK-----QLYMKFGIASITVLCYL 373
OY 375 AVTSIPSVSNALNMRSEFSFIQSTIGYVALLI---STFHVLIY-GMKRAPEEYRYRTPP 430
Db 374 LVLSLGIYRK-----RHFEFLYTHIILALLFLFCCHQOHKFIENGKE----- 416
OY 431 NEVLALVLPISIVILGKIILFLPCISIRKLRIKKMGERSQF 470
Db 417 ----WIVVSLIHW-----LEKRLFRIMWLOF 439

```

```

RESULT 8
FREQ_YEAST STANDARD: PRT: 1769 AA.
AC P42945;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 200.0 kDa protein in G2F3-IME2 intergenic region.
GN YJL109C OR J0808.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RA STRAIN=S288C / FY1679;

```

RX MEDLINE-96090136; PubMed-7483851;  
 RA Rasmussen S.W.;  
 RT "A 37.5 kb region of yeast chromosome X includes the SWE1, MEF2, GSH1  
 RT and GSD3 genes, a YCP-1-related gene, an open reading frame similar  
 RT to the DAL80 gene, and a tRNA(Arg).";  
 RL Yeast 11:873-883(1995).  
 CC -1- SIMILARITY: BELONGS TO THE BAP28 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HEAT REPEAT.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X85021; CA59385.1; -;  
 DR EMBL: Z49384; CA89404.1; -;  
 DR SGD: S0003645; YJL109C.  
 DR InterPro: IPR000357; HEAT\_repeat.  
 DR Prosite: PS50077; HEAT\_REPEAT; 1.  
 KW Hypothetical protein.  
 FT REPEAT 1729 1767 HEAT.  
 SQ SEQUENCE 1769 AA; 20080 MW; 064480DD1249B241 CRC64;

Query Match 4.28; Score 107; DB 1; Length 1769;  
 Best Local Similarity 19.94; Pred. No. 9.4; Indels 144; Gaps 22;

Matches 106; Conservative 79; Mismatches 205; Indels 144; Gaps 22;  
 QY 17 CLPFGINGIDARKVT-VGVIGSGDFAKSLTIRLIRCGYHVIGSRNPKASEFPPHVD 75  
 DB 163 CLSNFVSEKPPPLATMIKLFNDMDFLKLYTSYDQCIKH-----NATYNO----- 209  
 QY 76 VTHHEDALTNTNIFVAVHREHTSLMDLKRHLVGLKILIDVSNMKNRNOFP-----ESNAE 131  
 DB 210 -----LFTTCCFIV-----VAFNSNDEKKNQVLPILLETSK 244  
 QY 132 YLASLPD-SLIYGVNVSAMALQPKDASROYIICSNIQAROOVIELAROLNTP 190  
 DB 245 LLSKSKDCQIAHTLVFATLPLKTTIIMAMETILSLDQKE-----AKSALLTI 299  
 QY 191 D--LGSLSAREIENPLRLFTMRGPVVAISLATF-----FLYSEYRDVIHP 238  
 DB 300 CKLFQFLKCGGNVDQLSKIFLFDK-PDTVSILTFLEDKKPVCDKFTSYRSIAR- 357  
 QY 239 YARNQ-----OSDFKPIEIVNKTLPVATITLSLVLAGLAAAYOLYGTKRPRPPW 294  
 DB 358 YDMSKLTITLSLKKIKLEERE-----VRLITDLYLSELEEDKSOL-----VELEYF 407  
 QY 295 L-----ETWLCOROLGILSFEFFAM-----VHVAYSICLPMRSERYLFLNMA 337  
 DB 408 ISINEDLVKCLSKSLGTLGELFEIRLTSLEFNADVNTDIYKQSLDVEVETTKDTASFO 467  
 QY 338 YQOVHANIENTSMN-----EEFVWRIENTYSIFGINSIG-----LLSL 374  
 DB 468 FLUKHSELINTTVMMLTEGERKVKYLSLFEATIGKGYKASSPLTSFTTLESRIFFL 527  
 QY 375 AVTSIPSVNALMMRRESFQSTLGYALLISTFHVILYGMKRAFEERYRFTIPPEVL 434  
 DB 528 RVTISPAAPALK-----LISLNNAKY--INSIEKEVNIFFLVPLCLIC 569  
 QY 435 ALVLPISVIIGKIIILPCLISRLKRIKGMESQFLEEGIG-----TTPHVSPE 485  
 DB 570 ALMDASIKVATGVKILSLIA-----KRSTKHVFLSDKLYGENTYIPLNKR 617

RESULT 9  
 FSHR\_HUMAN  
 ID FSHR\_HUMAN STANDARD; PRT; 695 AA.  
 AC P23945;  
 DT 01-MAR-1992 (Rel. 21, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Follicle stimulating hormone receptor precursor (FSH-R) (Fol11tropin  
 DE receptor).  
 GN FSHR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE-93246012; PubMed-1301382;  
 RX MEDLINE-91222171; PubMed-1709010;  
 RA Muegish T., Nakamura K., Takakura Y., Itoh Y., Igarashi M.;  
 RT "Cloning and sequencing of human FSH receptor cDNA";  
 RL Biochem. Biophys. Res. Commun. 175:1125-1130(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Testis;  
 RX MEDLINE-93246012; PubMed-1301382;  
 RA Kelson C.A., Cheng S.V., Nugent N.P., Schweichardt R.L.,  
 RA Rosenthal J.L., Overton S.A., Wands G.D., Kuzaja J.B., Luchette C.A.,  
 RA Chappel S.C.;  
 RT "The cloning of the human follicle stimulating hormone receptor and  
 RT its expression in COS-7, CHO, and Y-1 cells";  
 RL Mol. Cell. Endocrinol. 89:141-151(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Tilly L.T., Alhara T., Nishimori K., Jai X.-C., Billig H.,  
 RA Kowalski K.I., Perlas E.A., Hsueh A.J.;  
 RT Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RP SEQUENCE OF 1-342 FROM N.A.  
 RC TISSUE-Testis;  
 RX MEDLINE-93075197; PubMed-1359889;  
 RA Gromoll J., Gudermann T., Nieschlag E.;  
 RT "Molecular cloning of a truncated isoform of the human follicle  
 RT stimulating hormone receptor";  
 RL Biochem. Biophys. Res. Commun. 188:1077-1083(1992).  
 RN [5]  
 RP SEQUENCE OF 1-51 FROM N.A.  
 RX MEDLINE-95011044; PubMed-7926278;  
 RA Gromoll J., Dankbar B., Gudermann T.;  
 RT "Characterization of the 5' flanking region of the human follicle-  
 RT stimulating hormone receptor gene";  
 RL Mol. Cell. Endocrinol. 102:93-102(1994).  
 RN [6]  
 RP 3D-STRUCTURE MODELING OF 49-228  
 RX MEDLINE-96363672; PubMed-8747461;  
 RA Jiang X., Dreano M., Buckler D.R., Cheng S., Ythier A., Wu H.,  
 RA Hendrickson W.A., el Tayar N.;  
 RT "Structural predictions for the ligand-binding region of glycoprotein  
 RT hormone receptors and the nature of hormone-receptor interactions";  
 RL Structure 3:1341-1353(1995)  
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY  
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE  
 CC ADENYLATE CYCLASE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: SEROTOLI CELLS AND OVARIAN GRANULOSA CELLS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC FSH/LSH/TSH SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```
DR EMBL: M65085; AA52477.1; -
DR EMBL: S59900; AA52480.1; -
DR EMBL: M95489; AA52478.1; -
DR EMBL: X68044; CAA48179.1; -
DR EMBL: S73199; AAB32071.1; -
DR PIR: JN0122; JN0122.
DR PDB: 1XUN; 15-MAY-97.
DR GCRDB: GCR_0071; -
DR GCRDB: GCR_0404; -
DR GCRDB: GCR_0588; -
DR GCRDB: GCR_0690; -
DR MIM: 136435; -
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR00372; LRR_Nterm.
DR Pfam: PF00001; 7tm_1; 1.
DR Pfam: PF00560; LRR; 4.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PR00373; GLYCHROMONER.
DR PRINTS: PR01143; FSHRECEPTOR.
DR SMART: SM00013; LRRNT; 1.
DR PROSITE: PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE: PS0262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation; Repeat; Leucine-rich repeat; Alternative splicing;
KW 3D-structure.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 695 FOLICLE STIMULATING HORMONE RECEPTOR.
FT DOMAIN 18 366 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 367 387 1 (POTENTIAL).
FT DOMAIN 388 398 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 399 421 2 (POTENTIAL).
FT DOMAIN 422 443 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 444 465 3 (POTENTIAL).
FT DOMAIN 466 485 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 486 508 4 (POTENTIAL).
FT DOMAIN 509 528 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 529 550 5 (POTENTIAL).
FT DOMAIN 551 573 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 574 597 6 (POTENTIAL).
FT DOMAIN 598 608 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 609 630 7 (POTENTIAL).
FT DOMAIN 631 695 CYTOPLASMIC (POTENTIAL).
FT REPEAT 69 93 LRR 1.
FT REPEAT 119 143 LRR 2.
FT REPEAT 170 192 LRR 3.
FT REPEAT 193 216 LRR 4.
FT REPEAT 218 240 LRR 5.
FT DISULFID 442 517 BY SIMILARITY.
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 285 MISSING (IN SHORT ISOFORM).
FT VARSPPLIC 342 695 MISSING (IN SHORT ISOFORM).
FT CONFLICT 13 13 S -> R (IN REF. 4).
FT CONFLICT 112 112 N -> T (IN REF. 1).
FT CONFLICT 197 198 EL -> AV (IN REF. 1).
FT CONFLICT 295 295 S -> P (IN REF. 4).
FT CONFLICT 307 307 T -> A (IN REF. 1).
FT CONFLICT 307 307 N -> S (IN REF. 1).
FT CONFLICT 680 680
SO -SEQUENCE 695 AA; 78294 MM; 723B8E71F76D2CD5 CRC64;
```

```
Query Match 4.2%; Score 106; DB 1; Length 695;
Best Local Similarity 16.4%; Pred. No. 3.8;
Matches 84; Conservative 81; Mismatches 148; Indels 198; Gaps 18;
```

```
OY 79 HEDALKTNIIVAIHREHYTSLSMDRLHVLG-----KIILDSNNMR 121
DB 98 HIRIEKANNL-LYINPEFOMLPNIQYLISNTGIKHLPDVHKIHSLOKVLIDODINT 156
OY 122 INQYESNAEVLASLEPDSLIYKGFNVASAMALQGPDKASROYICSNINIAQOVIEL 181
```

```
DB 157 IH-----TIERNSPVGLSFESVILML-----NKNGIOEIHNCA----- 189
OY 182 ARQNFIPIDGLSSARELENLPLRLFTLMKGPVVAISLATFFLYKSPVRYHPYAR 241
DB 190 ---FNGTQLEINLSDNNNIEELPNDVFGASGPVILDISRRIRHSLPSYGLENLKRLA 246
OY 242 NQOSDFYKPIELVNTKL-PIVATILSLVY----- 271
DB 247 RSTYNLKLPL-----TLKLVALMEASLTPYSHCCAFANMRQISELHPICKSLROE 300
OY 272 -----LAGLAAAYOLYGYTKYRRP-----PMLLETWLOCRKOL 305
DB 301 VDMQTRGQRSLAEDNSSYSGPDMYTEDEDYDLCEVVDYWCSPDPANPCEDIM 360
OY 306 G-----LSFFPAMVHA-----YSLCIPMRSEYLYFLNAYQOVHANIEEN 347
DB 361 GYNILRLVLMFISILAITGNITVLVLTTSQYKLTVP-----RLMCLNLAFAAD----- 408
OY 348 SNNEEVRHLEMYISFGISLGLSLAATSVSPVSNALNMR-----PSFIQSTL 398
DB 409 -----LCIGITVLLIASVDITHTKSOYHNVAIDMOTGACDAAGFTVPASEL 455
OY 399 G---YVALLISTFH-----VLITGMRKAFEEYRYPPNFV-LA 435
DB 456 SVYTLVAITLERHWTTHAMQLDCKYQLRHAASVVMGMTEFAALPLPIFGISSYMKVS 515
OY 436 LVLP-----STVILGKITLFLPC 453
DB 516 ICLPMDIDSPLSQLYVMSLVLNVLAFVVIC 546
RESULT 10
SOTB_ECO57
ID SOTB_ECO57 STANDARD; PRT; 396 AA.
AC P58529;
DT 01-MAR-2002 (rel. 41, Created)
DT 01-MAR-2002 (rel. 41, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Sugar efflux transporter.
GN SOTB OR Z2173 OR ECS2135.
OS Escherichia coli O157:H7.
OC Escherichia; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfal G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grodbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- FUNCTION: Involved in the efflux of sugars. The physiological role
CC may be the reduction of the intracellular concentration of toxic
CC sugars or sugar metabolites. Transports L-arabinose and to a
CC lesser extent IPTG. Seems to contribute to the control of the
CC arabinose regulon (By similarity).
```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS  
 CC THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE005352; AAG56236.1;  
 CC DR EMBL: AF002557; BAB35558.1;  
 CC KM TRANSPORT: Sugar transport; Transmembrane; Inner membrane.  
 CC FT TRANSMEM 15 35  
 CC FT TRANSMEM 50 70 POTENTIAL.  
 CC FT TRANSMEM 81 101 POTENTIAL.  
 CC FT TRANSMEM 103 123 POTENTIAL.  
 CC FT TRANSMEM 136 156 POTENTIAL.  
 CC FT TRANSMEM 170 190 POTENTIAL.  
 CC FT TRANSMEM 209 229 POTENTIAL.  
 CC FT TRANSMEM 246 266 POTENTIAL.  
 CC FT TRANSMEM 275 295 POTENTIAL.  
 CC FT TRANSMEM 319 339 POTENTIAL.  
 CC FT TRANSMEM 333 353 POTENTIAL.  
 CC FT TRANSMEM 364 384 POTENTIAL.  
 CC SQ SEQUENCE 396 AA; 42568 MW; CB6A209B0B6D4F0 CRC64;

Query Match 4.1%; Score 104; DB 1; Length 396;  
 Best Local Similarity 18.8%; Pred. No. 2.7;  
 Matches 85; Conservative 72; Mismatches 148; Indels 146; Gaps 20;

QY 31 VTVGVIGSGPFAKSLTRLRGCHVHVIGSRNPKFSEFPVVDVTHHEDALTKTNIIF 90  
 DB 31 VPGVLT--SDIAQSFHMQTQVQIMLTIVAVVVALMSIPEMLTMSOVERRKLCLFVFE 88  
 QY 91 VAIREHYTSIMDLRHLVGLIILDVSNMNRINQYPSNMEYLASLPDLSLVGFPNVVS 150  
 DB 89 IASIVLSFLS-WSTFVTVLISRI-----GVAFAHAIF-----WSTIRA 123  
 QY 151 AMAALQCP--KDSROYVYCSNNIQAQOYELAR-----QLNFPIDIGSL----- 195  
 DB 124 SLATRMAPAGRAQALSLIATGTAIAWLGPLGRIVGYQFWMETFPALIGALITLILC 183  
 QY 196 -----SSAREINLPLRLFTLMRGVVAISLAT-----FFLYSFVADVHPY 239  
 DB 184 LILKLLPLPSEHSGSLSLP-----LFRPALMSIYLLTVVVVTAHTTAVSY-----IEPF 235  
 QY 240 ARN---QOSDFYKIPRIEIVNKTLPVATLTSLVYLAGLLAAAYQLYYGFYRRFPWLE 296  
 DB 236 VQNTAGSANP-----ATALLLLLGAGCITIGSVFCKLIGNYAS----- 274  
 QY 297 TWIQCROQLGLSFFPAMVHVASLCPMRRSERYLEFLNMAVQOVHANINENSMNEEYWR 356  
 DB 275 -----ALVSTAIALLVCLALLLPANSE-----IHGVALSIF-----WG 309  
 QY 357 IEMVIRFGISGLSL-----LAVTSIPSVN-----ALNNRESFIQST 397  
 DB 310 IAMMIIIGLQGVKVALAPATDVAMALFSGIENFIIIGAGALVGNVSLHWS-----MSM 364  
 QY 398 LGYVALISTFHVLIYG-----WKRAFEER 422  
 DB 365 IGYVG-TVPAPALIMSIILFRMPVLEED 394

DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Sugar efflux transporter.  
 GN SOTB OR B1528.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RT [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,  
 RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [3]  
 RP SEQUENCE OF 217-396 FROM N.A.  
 RX MEDLINE=93186717; PubMed=8383113;  
 RA Cohen S.P., Haechler H., Levy S.B.,  
 RT "Genetic and functional analysis of the multiple antibiotic  
 RT resistance (mar) locus in Escherichia coli.";  
 RL J. Bacteriol. 175:1484-1492(1993).  
 RN [4]  
 RP CHARACTERIZATION.  
 RC STRAIN=SR0;  
 RX MEDLINE=99194728; PubMed=10094697;  
 RA Bost S., Silva F., Belin D.;  
 RT "Transcriptional activation of ydeA, which encodes a member of the  
 RT major facilitator superfamily, interferes with arabinose accumulation  
 RT and induction of the Escherichia coli arabinose PBAD promoter.";  
 RL J. Bacteriol. 181:2185-2191(1999).  
 RN [5]  
 RP CHARACTERIZATION.  
 RC STRAIN=JS219;  
 RX MEDLINE=99369894; PubMed=10438792;  
 RA Carole S., Pichoff S., Bouché J.-P.;  
 RT "Escherichia coli gene ydeA encodes a major facilitator pump which  
 RT exports L-arabinose and isopropyl-beta-D-thiogalactopyranoside.";  
 RL J. Bacteriol. 181:5123-5125(1999).  
 CC -1- FUNCTION: Involved in the efflux of sugars. The physiological role  
 CC may be the reduction of the intracellular concentration of toxic  
 CC sugars or sugar metabolites. Transports L-arabinose and to a  
 CC lesser extent IPTG. Seems to contribute to the control of the  
 CC arabinose regulation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS  
 CC THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

DE cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate  
DE cotransporter 1) (Renal Na(+)-dependent phosphate cotransporter  
1).



```

QY 303 KOLGLISFEFAMVHVASICLPMRSERYL-----FLNNAVQOVHANIENSME--- 351
DB 143 -----FREDDITINLCSTNLFYTLISACCALDPLMMY-----TNDDRK 184
QY 352 -----EEVWRLEMTISFIMISGLISLAVISIPVS-----NALWK 388
DB 185 INVINDINSOLKVEQHSHLMAVSLNSLLQLFVTSIVAFISFVLANMSMTNDELGNANLRLW 244
QY 389 -----REFSFIOST--LGYVALLISFPHVLYGMRKAFEEERYR---FTTPRFVL- 434
DB 245 FFAFKVLYRVSFHHSTHOLFMSTIYADFHAIISYGELOAI--SLRVSVFYDDPVMINE 302
QY 435 -----ALVPSIY-----ILKIIILFPCISRKLRKIK--GWEKSP- 470
DB 303 TGECAITWYGAKKMLMSFIQTALTIRIIRIIPVY-----KIKPKRQGLFNALRA 358
QY 471 LEEGIGGT 479
DB 359 LTAGISTTI 367

RESULT 15
FSHR_PIG
ID FSHR_PIG STANDARD: PRT; 695 AA.
AC P49059; 077514;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Follicle stimulating hormone receptor precursor (FSH-R) (Follictrophin
DE receptor).
GN FSHR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RX MEDLINE=96011644; PubMed=7590277;
RC TISSUE-Ovary;
RN SEQUENCE FROM N.A.
RA Remy J.J., Labhib-Mansais Y., Yerie M., Bozon V., Couture L.,
RA Pajot E., Grebert D., Salesse R.;
RT "The porcine follictrophin receptor: cDNA cloning, functional
RL Gene 163:257-261(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA Wang Y.F., Meyer K.B., Schmidt K., Wan S.J., Degen S.J.F.,
RA la Barbera A.R.;
RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RL "- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC ADENYLATE CYCLASE.
CC "- SUBCELLULAR LOCATION: Integral membrane protein.
CC "- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC FSH/LSH/TSH SUBFAMILY.
CC "- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L31966; AAA86933.1; -.
CC EMBL: AF025377; AAC24981.1; -.
CC HSSP: P23945; 1XUN.
CC GCRDB: GCR_1561; -.
CC InterPro: IPR000276; GPCR_Rhodpsn.
CC InterPro: IPR001611; LRR.

```

```

DR InterPro: IPR000372; LRR_Nterm.
DR Pfam: PF00001; 7tm_1; 1.
DR Pfam: PF00560; LRR_4; 4.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PRO0373; GLYCHROMONER.
DR PRINTS: PRO1143; FSHRECEPTOR.
DR SMART: SM00013; LRRNT; 1.
DR PROSITE: PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW phosphorylation; Repeat; Leucine-rich repeat.
FT SIGNAL 1
FT CHAIN 18 695
FT DOMAIN 18 366
FT TRANSMEM 367 387
FT DOMAIN 367 387
FT TRANSMEM 388 398
FT DOMAIN 399 421
FT TRANSMEM 422 443
FT TRANSMEM 444 465
FT DOMAIN 466 485
FT TRANSMEM 486 508
FT DOMAIN 509 528
FT TRANSMEM 529 550
FT DOMAIN 551 573
FT TRANSMEM 574 597
FT DOMAIN 598 608
FT TRANSMEM 609 630
FT DOMAIN 631 695
FT REPEAT 631 68
FT REPEAT 69 93
FT REPEAT 119 143
FT REPEAT 170 192
FT REPEAT 193 216
FT REPEAT 218 240
FT DISULFID 442 517
FT CARBOHYD 191 191
FT CARBOHYD 199 199
FT CARBOHYD 293 293
FT CONFLICT 2 2
FT CONFLICT 13 13
FT CONFLICT 60 60
FT CONFLICT 166 166
FT CONFLICT 215 215
FT CONFLICT 247 247
FT CONFLICT 257 257
FT CONFLICT 334 334
FT CONFLICT 349 349
FT CONFLICT 352 352
FT CONFLICT 383 383
FT CONFLICT 407 407
FT CONFLICT 421 421
FT CONFLICT 427 427
FT CONFLICT 435 435
FT CONFLICT 483 483
FT CONFLICT 550 550
FT CONFLICT 586 586
FT CONFLICT 607 607
FT CONFLICT 691 691
SQ SEQUENCE 695 AA; 78172 MW; E9EBDB29C79C450 CRC64;

Query Match 4.1%; Score 103; DB 1; Length 695;
Best Local Similarity 16.0%; Pred. No. 6.1; Indels 212; Gaps 22;
Matches 91; Conservative 95; Mismatches 169;
18 LPNG-INGIKDARKVTVGVIGSGDPAKSLTIRLCRGVHVIGSRNPKRASEFFPHVVDV 76
DB 61 IRKGAFSRGDLEKI-----EISQNDVLEVTIAN--VFSNPKL----- 97
QY 77 THHEDALTKNTNIIFAVIAIHEHYTSLMDLRHLVG-----KILDIVSN 119
DB 98 --HEIRIERANNUL-LYIDPDARCONLPIRLYLISNTGVKHLPAVHKIOSLOKVLDDIODN 154

```







Db 64 TAGLPPSLAQVTFQBEAVSSPEVIFVAVFRHHYSSLCSLADOLAGKILVDVSNPTKEERL 123  
Qy 124 QYPSNMEYLAFLPPDLSLVKGFNVASAMALQLPKDSAROVYICSNINQAROOVIELAR 183  
Db 124 QHRQSNMEYLAFLPPDLSLVKGFNVASAMALQACPROGNROVYICGDLLEAKHRYSEAR 183  
Qy 184 QLANFIPIDGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYARNO 243  
Db 184 AMGTPLDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDOVLPYIRKD 243  
Qy 244 OSDFFYKPIELFVNKTPRIVATITLSLVYLAGLAAAOQLYGTYKRRPPELMTWLOCRK 303  
Db 244 ENKFFKAPLPSVNTTIPCVAILLPGVLAALQLRGRTYQRRPDMLDHMLDHRK 303  
Qy 304 QLGILSFEFFAMVAVSYCLPMRSEERYLFNMAVQOQVHANIEANSNMEEWRIEMYSF 363  
Db 304 QIGLISFEFFAMVAVSYCLPMRSHRDYLVNLAVKOVLANKSRMLWEEVWRMEIYVSL 363  
Qy 364 GIMSLGSLSLAVTSIPSVSNAALMWRSEFIOSTLGYALLISFPHVLYGKRAFEERY 423  
Db 364 GVLAMGLSLAVTSIPSISANSIANKWKEFSVOSTLGEFVALLSTMHTLTGTWTRAFEBNH 423  
Qy 424 YRFYTPPFVAVLALPVSIVILGKIILFLPCISRLKRIKKGEMK 467  
Db 424 YKFTLPPTFTLLPVCYIILAKGLFLPLCLSHRLTKIRGWER 467

RESULT 2  
ID 092421 PRELIMINARY; PRT: 514 AA.  
AC 092421:  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE DUDULIN 2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Serru V., Lambdin D., Lenoir C., Manivet P., Vaudourdolle M.,  
RA Kellermann O., Loric S.;  
RT "Molecular cloning and expression of mouse dudulin 2."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY029586; AAK50539.1; -  
SQ SEQUENCE 514 AA; 57268 MW; 3398866C28BAEC02 CRC64;

Query Match 54.5%; Score 1382; DB 11; Length 514;  
Best Local Similarity 56.0%; Pred. No. 7.4e-101;  
Matches 260; Conservative 87; Mismatches 105; Indels 12; Gaps 4;

Qy 6 MMGSPKSLSETCPLNGINGIDARKVTVGVYSGDFAKSLIRLIRCGYHVYIGSRNPF 65  
Db 14 LVSDSGSLAE--VP-----EAPK--VGLISGDFARSLATRLVSGSFVYVSGSRNPKR 63  
Qy 66 ASEFFPVAVDVTTHEDALTNTNIFVAIHREHYSLMDLRLLVGLKILIDVSNMNRIN-- 123  
Db 64 TAGLPPSLAQVTFQBEAVSSPEVIFVAVFRHHYSSLCSLADOLAGKILVDVSNPTKEERL 123  
Qy 124 QYPSNMEYLAFLPPDLSLVKGFNVASAMALQLPKDSAROVYICSNINQAROOVIELAR 183  
Db 124 QHRQSNMEYLAFLPPDLSLVKGFNVASAMALQACPROGNROVYICGDLLEAKHRYSEAR 183  
Qy 184 QLANFIPIDGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYARNO 243  
Db 184 AMGTPLDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDOVLPYIRKD 243  
Qy 244 OSDFFYKPIELFVNKTPRIVATITLSLVYLAGLAAAOQLYGTYKRRPPELMTWLOCRK 303  
Db 244 ENKFFKAPLPSVNTTIPCVAILLPGVLAALQLRGRTYQRRPDMLDHMLDHRK 303

Qy 304 QLGILSFEFFAMVAVSYCLPMRSEERYLFNMAVQOQVHANIEANSNMEEWRIEMYSF 363  
Db 304 QIGLISFEFFAMVAVSYCLPMRSHRDYLVNLAVKOVLANKSRMLWEEVWRMEIYVSL 363  
Qy 364 GIMSLGSLSLAVTSIPSVSNAALMWRSEFIOSTLGYALLISFPHVLYGKRAFEERY 423  
Db 364 GVLAMGLSLAVTSIPSISANSIANKWKEFSVOSTLGEFVALLSTMHTLTGTWTRAFEBNH 423  
Qy 424 YRFYTPPFVAVLALPVSIVILGKIILFLPCISRLKRIKKGEMK 467  
Db 424 YKFTLPPTFTLLPVCYIILAKGLFLPLCLSHRLTKIRGWER 467

RESULT 3  
ID 09NVB5 PRELIMINARY; PRT: 488 AA.  
AC 09NVB5:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CDNA FLJ10829 FIS, CLONE NT2RP4001138 (DUDULIN 2).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y.,  
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Serru V., Manivet P., Lenoir C., Eschwege P., Lambdin D.,  
RA Vaudourdolle M., Kellermann O., Loric S.;  
RT "Dudulin 2, a new tumor antigen expressed in various human tumors."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK001691; BAA91839.1; -  
DR EMBL: AY029585; AAK50538.1; -  
DR InterPro: IPR003006; IG\_MHC.  
DR PROSITE: PS00290; IG\_MHC; UNKNOWN.1.  
SQ SEQUENCE 488 AA; 54616 MW; BC0BCA48335AAD6 CRC64;

Query Match 53.5%; Score 1356; DB 4; Length 488;  
Best Local Similarity 53.1%; Pred. No. 7.8e-99;  
Matches 260; Conservative 91; Mismatches 121; Indels 18; Gaps 6;

Qy 3 SISMMGSPKSLSETCPLNGINGIDARKVTVGVYSGDFAKSLIRLIRCGYHVYIGSRN 62  
Db 11 SLHVSDSGSLAE--VP-----EAPK--VGLISGDFARSLATRLVSGSFVYVSGSRN 60  
Qy 63 KFASEFFPVAVDVTTHEDALTNTNIFVAIHREHYSLMDLRLLVGLKILIDVSN--NM 120  
Db 61 PKRTARLPYSAQVTFQBEAVSSPEVIFVAVFRHHYSSLCSLADOLAGKILVDVSNPTKE 120  
Qy 121 RINQYPSNMEYLAFLPPDLSLVKGFNVASAMALQLPKDSAROVYICSNINQAROOVIE 180  
Db 121 EHLQHRQSNMEYLAFLPPDLSLVKGFNVASAMALQACPROGNROVYICGDLLEAKHRY 180  
Qy 181 LARLANFIPIDGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYA 240  
Db 181 MALAMGFMYPDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDOVLPY 240  
Qy 241 RNQOSDFYKPIELFVNKTPRIVATITLSLVYLAGLAAAOQLYGTYKRRPPELMTWLO 300  
Db 241 QESQNKKEFKPLPSVNTTIPCVAILLPGVLAALQLRGRTYQRRPDMLDHMLDHRK 300  
Qy 301 CRKQGLSFEFFAMVAVSYCLPMRSEERYLFNMAVQOQVHANIEANSNMEEWRIEMYS 360  
Db 301 HRKQIGLISFEFFAMVAVSYCLPMRSHRDYLVNLAVKOVLANKSRMLWEEVWRMEIY 360

```

Oy 361 ISFGMSIGLLSLAVTSPISVSNLNNRSEFSIOSTGCVALLSTTFPVLIYGMKRAE 420
Db 361 LSLGLVALGATISLNLAVTSLPSTANLNRRSEFSVOSLCEVALVSTLTLLTYGWRATE 420
Oy 421 EBYRYRTPPNFVNLVLPISIVILGKRIILFLPCISIKRLKRIKMGESQFLEEGIGCTP 480
Db 421 ESRKRYFLPPFTLLVPCVYILAKALFLPCISIRLARIRGMWR----ESTIKFTLP 476
Oy 481 --HVSPEKVT 488
Db 477 TDHALAEKTS 486

RESULT 4
0923B6 PRELIMINARY; PRT; 470 AA.
AC 0923B6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DS SIMILAR TO HYPOTHETICAL PROTEIN FLJ21353.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RS EMBL, BC006651, AAH06651.1;
SQ SEQUENCE 470 AA; 52994 MW; 6823E7682AC78B19 CRC64;

```

ID	NAME	PRELIMINARY	PT	AA
AD	O91W31			
AD	O91W31			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	TRANSMEMBRANE PROTEIN TIARP.			
GN	TIARP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Soturognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SWISS; TISSUE=ADIPOSE TISSUE;			
RX	MEDLINE=11424005; PubMed=11443137;			
RA	Moldes M., Lasnier F., Gauthereau X., Klein C., Patraut J., Fève B.,			
RA	Chambaut-Guerin A.M.;			
RT	"Tumor necrosis factor-alpha-induced Adipose-related protein (TIARP),			
RT	a cell-surface protein that is highly induced by tumor necrosis			
RT	factor-alpha and adipose conversion.";			
RL	J. Biol. Chem. 276:33938-33946(2001).			
RQ	EMBL; AJ319746; CAC41351.1; -			
RQ	SEQUENCE 470 AA; 52971 MW; F19239188E08663E CRC64;			

Query Match	44.8%;	Score 1134;	DB 11;	Length 470;
Best Local Similarity	47.5%;	Pred. No. 2.4e-81;		
Matches 212;	Conservative 90;	Mismatches 142;	Indels 2;	Gaps 2

QY 33 VGVIGSGDFAKSLTIRLCRGYHVIVISRNPKFASSEFFPHVDVYTHHEDALTKTNIFVA 92  
| : |||| | | : : : | | : | | | : | : : | : : : | : |  
Db 22 VCJFTGTDEKSKLGLMLQGYSIVFGSRNPQ-VSSLPRGAELYSYSEAKSKSDIITLA 80

```
93 IREHYTSLMDLRLHLVGKILIDVSNMNRKIQDPESNAEYIASLPDLSLYKGENVSAW 132 y
      : : : : : : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 MHEHSDSLTELVDCLGKRVLVDSNNRKINQDPESNAEYLQLEPGAHVYKAINTISAW 140 b
```

```

QY 153 ALDGPKDSRQVYICSNINIOARQOVIELARQINFIPIDEGISSAREIENDPURTJTM 212
    ||| | |||||::| : : : : ||| : | ||| |||::||| :|
Db 141 ALGGSTLDASRQVFCGNDSKAKHRVMDIARTGLTPIEDGSLMAASEIENYPLQLEPMM 200

```

**Oy**    213 RGPVVVAISLATEFFELYSFVDVYIHPARNQSDFKPIEIVKTLPIAITLLSVYL     272  
     | : : | ||| : ::||| :     | : | : | : | : |||  
**Db**    201 RFPELVSSVLCTFEFYCAIREVIYPVNGKTDATYRLAISIPNRFPTALILLAVYL     260

QY 273 AGLLAAAYQLYYGTYKRRFPWLETWLQCRKQLGLSTFEFAMVHVAYSLCLPMRSERYL 332  
| : ||| ||| ||||| | : : ||||| : | : || : | : | :  
Db 261 PGILAIITLQIRGTYYRRFPWWLDHWMLCRKQLGLVALGSAFLHVITYTLVIPDIRYYWR 320

QY 333 FLNMAVQGVHANIEINSWNEEVRRIEMAYISFGINSLGLTSLAVTSIPSVSNALNRREFS 392

DG 321 LRAATITQALTNKSDPEITSYAMINDSYLAIGLIGFELILGTSTSPSVSNMVMNRER 380

QY 393 FIOSTLGVALLISTFHVLIYGMKRAFEETTYRPFVLAIVLSIILGILFLP 452

453 CISRKLRIKKGWK-SQFLEEGIGG 477

RESULT 6

Db 441 CIDKTLTRIQGWERNISKYTQALNG 466

Q91ZE8  
ID Q91ZE8 PRELIMINARY; PRT; 474 AA.  
AC Q91ZE8;  
DN 01 DEC 2001 (REMOVED 10 Cited)

DE DUDULIN 4.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OY	361	ISFGMSGLSLAVTSIPSPSNALNNRSEFIOSTGCVALLSTFPVLIVGWKRAE	420
Db	361	LSLGVLAGTSLAVTSIPSTANSLNRRSFVSSDCEVALVSTLTFTLVGWTRAE	420
OY	421	EERYVFPPPNFVNLVLPISIVILCKITILFLPCISNRKLRIKKGWESQFLEEGIGCTP	480
Db	421	ESRYKFYLPPPTTLTVPCVVITLAKALEFLPCISRRLAIRRWGR-----ESTIKFTLP	476
OY	481	--HNSPERVT 488	
Db	477	TDHAALEKTS 486	
RESULT	4		
ID	0923B6	PRELIMINARY;	PRT; 470 AA.
AC	0923B6;		
DT	01-DEC-2001 (TREMBLrel, 19, Created)		
DR	01-DEC-2001 (TREMBLrel, 19, Last sequence update)		
DE	01-DEC-2001 (TREMBLrel, 19, last annotation update)		
DS	SIMILAR TO HYPOTHETICAL PROTEIN FLJ23153.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Stratberg R.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.		
SR	EMBL; BC006651; AA06651.1; -		
SO	SEQUENCE 470 AA; 52994 MW; 6823EF682AC78B19 CRC64;		
Query Match	45.0%; Score 1141; DB 11; Length 470;		
Best Local Similarity	48.0%; Pred. No. 6.7e-82;		
Matches 214; Conservative 89; Mismatches 141; Indels 2; Gaps 2;			
OY	33	VGVIGSGDFASLTRLIRCGYHAVYIGSRNPFASEFPFHVDVTHHEDALTKNTIIFA	92
Db	22	VCIFRTGPGSLGKMLCCGYTFGSKRNP-VSSLPRGAELVSYSFAASKSDITIITA	80
OY	93	IHRHYTSLMDLRHLVGKILIDVSNMNRINOPESNMEYIASLFPSDLIVKGFNVSAM	152
Db	81	MHREHYDSLTELVDYDKGKVLVDVSNMNRKINOPESNMEYIAQLDPGAHVVAFTISM	140
OY	153	ALDIGPKRASQOYICSNNOIQAROQVIELAROLFPIIDLGSLSARETENLPRLFTIW	212
Db	141	ALOSTGTIDASROVFCGNDKSRAKORVMQDIARTLGTPIDLOGSLMAASETENPLDLPMW	200
OY	213	RGPVVAISLATPFPLYSFVRDVHPVARNOQSDPYKIPIELVNKTIRPAVATLTIISVL	272
Db	201	RPFPLSVLCFFPFVYCAIREVIYIPVNGKDATFYRLAISTIPNNVPFTALIALVYL	260
OY	273	AGLLAAAYOLYYGTYKRFRPPMLETWLOCRKOGLSFEFFAMVHAAYSCLPMRSEREL	332
Db	261	PGILLAIQLIVRGKYRRFRPMWLDMMLCMCRQDGLVALGFARLAHYITYLYTIRYVRR	320
OY	333	FLNNAVQOVHANIEANSNEEEVYRIEMTISFGIMSLGISLTAIVSIDSVSNALMWREFS	392
Db	321	LRNAITIGALTNNKSDPFTSYAMINDSYALGILGFFLLGIISLPSNMVMWRPR	380
OY	393	FIOSTLGVALLISFENHLYIGMKRAFEEYREVTPNFNVATLVPISVILGKITLFLP	452
Db	381	FVQSRLGLVTLVLTAAHTLVYGGKRFSLPSILRMSSLPAYITALVIPCADVLCITLMP	440
OY	453	CISRLKRIKKGWK-SQFLEEGIGG 477	
Db	441	CIDKTLTRIQRGMERNISKYTQSALNG 466	
RESULT	5		
	91IW31		

Query Match	44.8%;	Score 1134;	DB 11;	Length 470;
Best Local Similarity	47.5%;	Pred. No. 2.4e-81;		
Matches 212;	Conservative 90;	Mismatches 142;	Indels 2;	Gaps 2

QY 33 VGVIGSGDFAKSLTIRLCRGYHVIVISRNPKFASSEFFPHVDVYTHHEDALTKTNIFVA 92  
| : |||| | | : : : | | : | | | : | : : : | | : |  
Db 22 VCJFGTGDEKSLGLMLQGYSIVESGRNPQ-VSSLPRGAELVSYSEAKSKSDIITLA 80

```
93 IREHYTSLMDLRLHLVKGILLIDVSNMKNINQYPSNAEYIASLPDSLYKGVNSVAW 132 y
   : : : : : : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 MHEHEDLSLELVDCIAGKVIVDVSNKKINQYPSNAEYLADPEGHHVYKANITISAN 140 b
```

```

QY 153 ALDGPKDSRQVYICSNINIDARQOVIELARQINFIPIDEGISSAREIENDPURTJTM 212
    ||| | |||||::| : : : : ||| : | ||| |||::| : |
Db 141 ALGGSTLDASRQVFCGNDSKAKHRVMDIARTGLTPIEDGSLMAASEIENYPLQLEPMM 200

```

QY 213 RGPVVVAISLATEFFELYSFVDVYIHPARNQSDFKPIEIVKTLPIAITLLSVYL 272  
+ + :: + |||:: : ||::|| : ||::|| : ||::||  
Db 201 RFPEYLSVLCIFFEVYCAIREVIYPVNGKTDATYRLAISIPNRFPTALILLVL 260

QY 273 AGLLAAAYOLYYGTYKRRFPWLETWLQCRKOGLSTFEFAMVHVAYSLCLPMRSERYL 332  
| : ||| ||| ||||| | : : ||||| : | : || : | : | :  
Db 261 PGILAIITLQIRGTYYRRFPWWLDHWMLCRKQLGVALLGFAFLHVITYTLVIPRIYYRW 320

QY 333 FLNMAVQGVHANIEINSWNEEVRRIEMAYISFGINSLGLTSLAVTSIPSVSNALNRREFS 392  
| : : | : : || : :||| : :||| : :|||  
Db 321 LRAATITQALTNKSDPEITSYAMINDSYLAIGLIGFELILGISTLSIPSVSNMVMNRER 380

QY 393 FIOSTLGVALLISTFHVLIYGMKRAFEETTYRPFVLAIVLSIILGILFLP 452

453 CISRKLRIKKGWK-SQFLEEGIGG 477

RESULT 6

Db 441 CIDKTLTRIQGWERNISKYTQALNG 466

Q91ZE8  
ID Q91ZE8 PRELIMINARY; PRT; 474 AA.  
AC Q91ZE8;  
DN 01 DEC 2001 (REMOVED 10 Cited)

DE DUDULIN 4.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP Serru V., Lambdin D., Manivet P., pernet P., Vaubourdolle M.,  
 RA Kellermann O., Loric S.,  
 RT "Molecular cloning and expression of two new members of the dudulin  
 family".  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY029778; AAK40270.1;  
 SO SEQUENCE 474 AA; 53543 MW; 937EBA3393D13956 CRC64;

Query Match 43.2%; Score 1094.5; DB 11; Length 474;  
 Best Local Similarity 47.8%; Pred. No. 3.2e-78;  
 Matches 208; Conservative 83; Mismatches 143; Indels 1; Gaps 1;

QY 33 VGVIGSDGDFKSLTIRLCRGYHIVIGSRNPKFSEFPFHVDVTHEDALTKTNIFVA 92  
 DB 22 VCFGTGDFGSKMGKMLQCGYSIVFGSRNPQ-VSSLLPRCAEVLVSYSEAKSKSDIILA 80  
 QY 93 IHRHRYTSLMDLRHLVGLKILIDVSNMRTINQYPSNAEYLLAFPSLIYKGFNVYSAW 152  
 DB 81 MHRHRYTSLMDLRHLVGLKILIDVSNMRTINQYPSNAEYLLAFPSLIYKGFNVYSAW 140  
 QY 153 ALQGPDAQRQYVYICNNIQAQOVIETLARIQNLFIPIIDIGSLSSAREIENTPLRFTLW 212  
 DB 141 ALQSGTLDASROYLVCGNDSKAKORVMDIARTLGLTFLDQSLMAASEIENYPLQLPWM 200  
 QY 213 RGPVVAISLATPEFLKSPFVDVHPRARNOQSDFKIPRIEIVNKTPIVAITLISLYL 272  
 DB 201 RFPYLSLVLCIFPEFLCAIREVYIPYMGNTDVTYRLAISTPKRVPPTALILALVYL 260  
 QY 273 AGLLAAVYQLYYGTGKRRFPFMEWTLQCRKQGLLSFFPAMVVAVSLCPMRSRSYL 332  
 DB 261 PGILAAILOLYRGKRYRPFMDHMLCKRQGLLALGAFALHAIYTVIPYVYRWR 320  
 QY 333 FLNMAVQOYHANIENSNNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALNREPS 392  
 DB 321 LRKATITQALTSKROPFTSTYAMINDSYALGILGFLLVLYGTTSPSVSNMNMREFR 380  
 QY 393 FIOSTIGVALLISTFPHVLIYGMKRAFEERYRFPNPFVLAIVLSIYILGIIIFLP 452  
 DB 381 FVQSKLQYLVLTCTANTLVYGGKRLSPSLKMSLPNAVILALVITCAVAVLCILMP 440  
 QY 453 CISRKLRIKRGWEK 467  
 DB 441 CIDKTLIRIQGLGK 455

RESULT 7  
 Q9GL50 PRELIMINARY; PRT; 338 AA.  
 AC 09GL50;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE SIX TRANSMEMBRANE ENDOTHELIAL ANTIGEN OF PAEC.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP Nagasaka T., Bouillon G., Coupel S., Coulon F., Tesson L.,  
 RA Heslan J.-M., Souillou J.-P., Charreau B.;  
 RT "Differential gene expression in endothelial cells during TNF-alpha-  
 and LPS-mediated activation".  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF319659; AAG33868.1;  
 KW Transmembrane.  
 SO SEQUENCE 338 AA; 39918 MW; ED490E86E067A32B CRC64;

Query Match 31.6%; Score 801; DB 6; Length 338;  
 Best Local Similarity 49.7%; Pred. No. 2.9e-55;  
 Matches 148; Conservative 64; Mismatches 78; Indels 8; Gaps 2;

QY 170 NNTARQOVIETLARIQNLFIPIDL-GISLSAREIENLRLFTLMRGVVAISLATPEFL 228  
 DB 34 NSMRKPMVLHLHQTAFHDEFDCPELQHKOE-----LEPKWHLPIKIAALVSSLTFL 86  
 QY 229 YSFVVDVHPRARNOQSDFKIPRIEIVNKTPIVAITLISLYLAGLLAAVYQLYGTYK 288  
 DB 87 YTLREVIHPRVNSHQOYFKIPILVINKVLPKMSITLLALVLPYIAALVQAHNTKTY 146  
 QY 289 RRPFWLETLQCRKQGLLSFFPAMVVAVSLCPMRSEERYELNMAVQOYHANIENS 348  
 DB 147 KRPFWMDRMVWTRKQFGLLSFFPAMVVAVSLCPMRSEERYELNMAVQOYHANIENS 206  
 QY 349 WNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALNREPSFIOSTIGVALLISTF 408  
 DB 207 WIEDHVRMEIYVSLGIVTALIALAVTSIPSVSDSLTWREFFHYIOSTIGVALLISTF 266  
 QY 409 HVLVYGMKRAFEERYRFPNPFVLAIVLSIYILGIIIFLPICISRKLRIKRGWE 466  
 DB 267 HALIFAMNKVNDIKQFTWTPPTMIAVFLPYVLYICKVILLPLCLRKILRIKRGWE 324

RESULT 8  
 Q92422 PRELIMINARY; PRT; 339 AA.  
 AC 092422;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE DDDULIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP Serru V., Manivet P., Lambdin D., Vaubourdolle M., Kellermann O.,  
 RA Loric S.;  
 RT "Prostate and non-prostate expression of dudulin, the mouse ortholog  
 of human STEAP-2".  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY029584; AAK50537.1;  
 SO SEQUENCE 339 AA; 39109 MW; 32A2C29FE2E33BD0 CRC64;

Query Match 30.3%; Score 768; DB 11; Length 339;  
 Best Local Similarity 54.7%; Pred. No. 1.2e-52;  
 Matches 141; Conservative 53; Mismatches 64; Indels 0; Gaps 0;

QY 209 FTLMRGPVVAISLATPEFLYSFVDVHPRARNOQSDFKIPRIEIVNKTPIVAITLIS 268  
 DB 68 FPMWRPLPVKVAATISSLTFTLREITLPLVTSKREYFKIPILVNTTLPCVAYVLLS 127  
 QY 269 LVYLGLLAAVYQLYGTYKRRFPFMEWTLQCRKQGLLSFFPAMVVAVSLCPMRSR 328  
 DB 128 LVYLPGVLAALDQRLRGTQYQRPDMDLHWLQHRKQGLLSFFPAMVVAVSLCPMRSR 167  
 QY 329 ERYLFLNMAVQOYHANIENSNNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALN 388  
 DB 188 HRYDLVNLAVKQVLANKEDEHVDHVRMEIYVSLGIVTALIALAVTSIPSVSDSLTW 247  
 QY 389 RFPFIOSTIGVALLISTFPHVLIYGMKRAFEERYRFPNPFVLAIVLSIYILGIIIFLP 448  
 DB 248 RHRHYIOSKLGIVSLGITHALVFAFNKKNVDSQFVWYHPTPTMIAVFLPYVLYICKIA 307  
 QY 449 LFLPCISRKLRIKRGWE 466  
 DB 308 LCLPCLARKTLKIRGWE 325

RESULT	9			
09CWM7	09CWM7	PRELIMINARY;	PRT;	339 AA.
AC	09CWM7;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	2410007B19RIK PROTEIN.			
GN	STEAP OR 2410007B19RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;			
KX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alzawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner T., Batilov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,			
RA	Schirral L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baidarrelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilmink L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontsuki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection";			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK010437; BAB6938.1; -			
DR	MGD: 1917608; Steap.			
QO	SEQUENCE 339 AA; 39264 MW; 3F7AB9C7520F0968 CRC64;			

Query Match	30.08%	Score 760:	DB 11,	Length 339;
Best Local Similarity	53.18%	Pred. No. 5e-52;		
Matches 137/	Conservative 55;	Mismatches 66;	Indels 0;	Gaps 0;
QY 209	FTLMKGPVVAVIAISLATEPFELYSFVADVTHIPAPARNQOSDFYKPIPIEIVNKTLPITVAITLLS	268		
Db 68	FPNNRLPYKVAIAISSLSLFEVLTLLREIITYPLVTSSEQYFKPIPIIVINKVLPMAVITLLA	127		
QY 269	LVYIAGLLAAAYQLYTGGYKRYRFPFWLETWLCQKROGLSFFEFMAHVAVSLCLPMRRS	328		
Db 128	LVYIPGELTAAVQVLNKGKRYKRPFWLDRMMLAKQOFILSFEFVAVLAAVYSLSPMRRS	187		
QY 329	ERYFLFNAAVQOVAVIANINSNWNEEVMRIEMVYISFGISMLGLSLTAAVTSIPSVSNALNM	388		
Db 188	YRYKLIMNAYQVOQNKDKDAVNDVHMVMEIYVLSIGYIGALALALAAVTSIPSVSDSLWM	247		
QY 389	KEEFSTIOSTLCTVALLISTFHVLIYGWKRFAEEERYRYTPPNFVALVLPISVILGKTI	448		
Db 248	REFPHIOSKLCIVSLTCTVHALVFAVMNKKVVDVSQFVWYMPETPEMIAVFLPLVLICKIA	307		
QY 449	LFLPCISRKTLRIKKGWE	466		
Db 308	LCFLPCLRKKILKIRGWE	325		
RESULT 10				
ID 0924J9				
AC 0924J9:	PRELIMINARY:			
DT 01-DEC-2001 (Tremblurel, 19, Created)	PRT: 339 AA.			
DT 01-DEC-2001 (Tremblurel, 19, Last sequence update)				

DT 01-DEC-2001 (Tremblarel, 19, last annotation update)  
DE SIX-TRANSMEMBRANE EPITHELIAL ANTIGEN OF THE PROSTATE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=21371909; PubMed=11479226;  
RA Yang D., Holt G.E., Velders M.P., Kwon E.D., Kast W.M.;  
RT "Murine six-transmembrane epithelial antigen of the prostate, prostate  
stem cell antigen, and prostate-specific membrane antigen: prostate-  
specific cell-surface antigen and prostate highly expressed in prostate cancer of  
transgenic adenocarcinoma mouse prostate mice.";  
RT Cancer Res. 61:5857-5860(2001).  
RL EMBL; AF297098; AAK83126.1; -.  
DR Transmembrane.  
KW AA; 39318 MW; 4B26A71FF559E84F CRC64;  
SEQUENCE

Query Match	29.9%	Score 757	DB 11	Length 339
Best Local Similarity	53.1%	Pred. No. 8.6e-52		
Matches 137	Conservative 55	Mismatches 66	Indels 0	Gaps 0
QY	209	FTLMGPPVVAISLSTFFEFVSVFVDVHIPVARNQSDPYKPIPIEIVNKLPIVAITLLS	268	
Db	68	FPNMRPLPKVAVAIISLFLPYLLLEIIPYLTSSNQGYKKPIPIVINKVLEPMVAITLLA	127	
QY	269	LVYIAGLLAAAYVQLYGYKKRRRPPWLETFMLQCRKQLGSLFFEFVAVVAVSLCLPMRRS	328	
Db	128	LVYIPEGLAAVAVQLENGRTKKKFPWMLDRMLARKQFGLSPFFAVLVAVSYLSVPMRRS	187	
QY	329	ERYLEFLNAYQOVANINENSNEEYWRLEMTISFGIMSLGILLSLATYSIPSVSNALNW	388	
Db	188	YRYKLNNAYQOVQONKEDAEVHEDVWRMEIYVSGIYGLATLALLAVYSIPSVSDSLTW	247	
QY	389	REFSISTGTCVALLISTFHVLIYIGWKRAPEEYRFTPNFVALVLPISIVILGKII	448	
Db	248	REFPHYSIKSLIVSLSTGTHALVFAVMKMKWDVQSOFWYMPPTFMIAVFLPYLVILCKIA	307	
QY	449	LFLPCISRKLRKKKGWE	466	
Db	308	LCLPCLRKILKIRCGWE	325	
RESULT	11			
Q9D5R1				
ID	Q9D5R1	PRELIMINARY	PRT	132 AA.
AC	Q9D5R1			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	4921538B17RK	PROTEIN.		
GN	4921538B17RK			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Salto T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Salto R.,			
RA	Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schiraldi L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Botfell D., Bojunga N., Carninci P., de Bonaldo M.F.,			

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Segs T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL: AK015015; BAB29672.1; -  
DR MGD: MGI:1921301; 4921538B17R1K.  
SQ SEQUENCE 132 AA; 14931 MW; B56F8732AEBEAEF3 CRC64;

Query Match 25.3%; Score 642; DB 11; Length 132;  
Best Local Similarity 93.2%; Pred. No. 3e-43; Mismatches 0; Gaps 0;  
Matches 123; Conservative 6; Indels 0;

QY 359 MYISFGIMSLGLSLAVTSIPSVSNALNMRFSFIQSTIGYVALLSTFHVLYGKRA 418  
DB 1 MYISFGIMSLGLSLAVTSIPSVSNALNMRFSFIQSTIGYVALLSTFHVLYGKRA 60  
QY 419 FEERYFYPPNPNVLAIVLPSTYILKILFLPCISRKLRKKGKESQFLERIGGT 478  
DB 61 FAEEYFYPPNPNVLAIVLPSTYILKILFLPCISRKLRKKGKESQFLERIGGT 120

QY 479 IPHVSPEPVYV 490  
DB 121 VPHISPEPVYV 132

RESULT 12  
Q9H5R1 PRELIMINARY; PRT; 283 AA.  
ID Q9H5R1;  
AC Q9H5R1;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE CDNA: FLJ23153 FIS, CLONE LING09441.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsunura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
RA Isogai T., Sugano S.;  
RT "MEDO human cDNA sequencing project."  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK026806; BAB15559.1;  
DR InterPro: IPR000585; Hemopexin.  
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN\_1.  
SQ SEQUENCE 283 AA; 31814 MW; DDD330DA7C607D06 CRC64;

Query Match 24.8%; Score 628; DB 4; Length 283;  
Best Local Similarity 43.2%; Pred. No. 1e-41;  
Matches 123; Conservative 56; Mismatches 80; Indels 26; Gaps 5;

QY 13 LSTENC---LNGINGIDARKVTVGSGDFAKSLIRLRGCVHYVTSRNPKEFSEF 69  
DB 1 MERTCIDALPLVMN--SSEKQETVCITGTDGFGSLKMLQCGYSVFGSRNQ-KTTL 57  
QY 70 FPHVVDTHEDALTKTNIIFVAIHREHYTSLMDLRHLGKILIDVSNMRLINOYPSN 129  
DB 58 LPGAGVLYLSSEAKKDDIIIIAHRHYDFELPTELVANGKILVDSNNLKNINOYPSN 117  
QY 130 AEYLAFLPDSLIVKGFNVSAWALDGPKDSROYICSNNOAROOVIELARQLNFI 189

||||| : : : : : ||||| : : : : : ||||| : : : : :  
DB 118 AEYLAHVPGAHVVKANNTISAWALDQALDASRQVFCGNDSSKAKORVDIVNGLTP 177  
QY 190 IDGSLSSAREIEVPLRLFTLMGRGPVVAISLAFPELYSPFVDVHPARNQSDPYK 249  
DB 178 MDQCSLMAAKEIEKYPDLQLEPMMRFPPYLSAVLCVLFEEFCVIRDYVPPYEEKDNFR 237  
QY 250 IPRIEIVKTLPIVAITLLSLVYLAGLAAVQLYGTRKVRFPW 294  
DB 238 MAISIPNRIPI-----TAVYTACFG-----LPPW 262

RESULT 13  
Q9Y6U5 PRELIMINARY; PRT; 264 AA.  
ID Q9Y6U5;  
AC Q9Y6U5;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE WOGSC:H\_RG087E15.1 PROTEIN (FRAGMENT).  
DE WOGSC:H\_RG087E15.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99063792; PubMed-9847074;  
RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome sequence."  
RL Genome Res. 8:1097-1108(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strong C., Layman D., Graves T., Stromatt C.;  
RT "The sequence of Homo sapiens BAC clone CFB-87E15."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC005061; AAD43182.1; -  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 264 AA; 31203 MW; 35C948300357E72 CRC64;

Query Match 22.8%; Score 577; DB 4; Length 264;  
Best Local Similarity 56.4%; Pred. No. 9.8e-38;  
Matches 106; Conservative 36; Mismatches 46; Indels 0; Gaps 0;

QY 208 LFTLMGRPVVAISLAFPELYSPFVDVHPARNQSDPYKIPRIEIVKTLPIVAITLL 267  
DB 77 LFPQWHLPIKIAVMSLFLYTLRLREVHPLATSHQOYFKIPILVINKVLPWVSTILL 136  
QY 268 SLVYLGLAAVQLYGTRKVRFPWLETLQCRKOLGLSPFPANVHAYSLCLPMR 327  
DB 137 ALVYLPVIAIYQVHNGTYYKKFPHWLDKMTLRKQFGLSLFPAVLAITYLSAMRR 196  
QY 328 SERVLEPMNAVYQVHNIENSWNEEWEVRIEMTISFGIMSLGLSLAVTSIPSVSNALN 387  
DB 197 SYRKLLNMAVYQVQOONKEDAWIEHDVWRMEITYLSGLVLAITLAVTSIPSVSDSL 256  
QY 388 WREFSFQ 395  
DB 257 WREFHYQ 264

RESULT 14  
Q9H7Y1 PRELIMINARY; PRT; 143 AA.  
ID Q9H7Y1;  
AC Q9H7Y1;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

```

DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE CDNA FLJ14101 FIS, CLONE MAMMA1000859.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Issagi T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagahtsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024163; BAB14842.1;
SQ SEQUENCE 143 AA; 16012 MW; DF84CA139545B9 CRC64;

Query Match 16.1%; Score 409; DB 4; Length 143;
Best Local Similarity 58.9%; Pred. No. 8e-25;
Matches 73; Conservative 25; Mismatches 22; Indels 4; Gaps 1;

QY 357 IEMVIFGIMSGLLSLAVTSIPSVSNALNMRPFIQSTLGVALLSTFHVLIYQWK 416
DB 1 MEYISLGVALLGTSLAVTSIPSIANSINLNRHFSFVQSSLSGVALVSTLHTLYGWT 60

QY 417 RAFEERYRFPNPNVLAIVLPISVILGKIIFLPCISIKRIRKKGWKSQFLEGI 476
DB 61 RAFEESRYKFTLPFTTLIVPCVILAKALLPLFCISRRLARIRGWER-----ESTIK 116

QY 477 GTP 480
DB 117 FTLP 120

RESULT 15
Q930K7 PRELIMINARY; PRT; 198 AA.
AC Q930K7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SMA0349.
GN SMA0349.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid PSYMA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Hultzer L., Hymen R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti PSYMA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007212; AK64846.1;
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 20324 MW; C893B5F30A6511B CRC64;

```

Query Match 6.3%; Score 160; DB 16; Length 198;  
 Best Local Similarity 26.7%; Pred. No. 5.3e-05;  
 Matches 55; Conservative 40; Mismatches 83; Indels 28; Gaps 8;

```

QY 32 TVGVIGSGDFPAKSLTITRLRCGYHVI-GSRNPKFAS---EFPFHVVDYTHHEDATKTN 87
DB 3 TVAITGAGALGSLAERFTAAQIPAITANRGRGASLSVYDRGASVKAKVELKDAL-QND 61

QY 88 IIFVAIHREHYTSLMDLRHLV--GKILIDVSNMNRINQY-----ESNAEYLASLFP 138
DB 62 VVILAV---PYDSIADIVTQVSDMGQIYVDASNAIDFPAFKPRDLGRSLSTEIVSELP 118

QY 139 DSLIVGPNVVSAMALQLGPK--DASROYICSNNIQARQVTELARQLNFIPIIDGSL 196
DB 119 GAKVVAEFNTLPAAVLAADPDKGTGSRVLELSGNHSDANRQVAVELLSSLSGFAVDLGTIA 178

QY 197 SAREIENLPLRLFTLRGPPVVAISL 222
DB 179 ASGPIDQF-----GRPLVALNL 195

```

Search completed: May 9, 2002, 03:22:15  
 Job time: 4212 sec

THIS PAGE BLANK (USPTO)

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 22:28:40 ; Search time 308.42 Seconds  
(without alignments)  
10526.831 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagctgagtgctgcg.....gtcaattatcgtgtgtga 1891

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

N.Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1608.8	85.1	4329	22	AA515801 Human ORF of Six-T
2	1534.8	81.2	1680	22	AA515802 Human cDNA encoding
3	1466.2	77.5	1725	22	AA515793 Human DNA for Six-
4	1409.8	74.6	2453	22	AA507072 Human six transmem
5	1369.2	72.4	1561	22	AA515811 Human cDNA encoding
6	1367.6	72.3	2238	22	AA515810 Human ORF2 of Six-
7	1216.8	64.3	2102	22	AA515813 Human cDNA encoding
8	1093.2	57.8	2192	23	AA576493 DNA encoding novel
9	1093.2	57.8	3900	23	AA564300 DNA encoding novel

10	530.6	28.1	1953	22	ABA45820 Human breast cell
11	530.6	28.1	1953	22	ABA56335 Human foetal liver
12	530.6	28.1	1953	22	ABA25976 Probe #4442 for ge
13	530.6	28.1	1953	22	AAK04514 Human brain expres
14	530.6	28.1	1953	22	AAK30016 Human bone marrow
15	530.6	28.1	1953	22	AAK14606 Probe #4539 for ge
16	530.6	28.1	1953	22	AAI35978 Probe #4664 used t
17	530.6	28.1	1953	22	AAI104420 Probe #4411 used t
18	528.6	28.0	539	22	ABA50938 Human breast cell
19	528.6	28.0	539	22	ABA68917 Human foetal liver
20	528.6	28.0	539	22	ABA35870 Probe #1436 for g
21	528.6	28.0	539	22	AAK17249 Human brain expres
22	528.6	28.0	539	22	AAK43038 Human bone marrow
23	528.6	28.0	539	22	AAI23801 Probe #13734 for g
24	528.6	28.0	539	22	AAI149113 Probe #17799 used
25	528.6	28.0	539	22	AAI09410 Probe #9401 used t
26	528	27.9	528	22	AA515797 Human Six-Transmem
27	528	27.9	528	22	AA515806 Human Six-Transmem
28	523.4	27.7	525	22	AA515796 Human Six-Transmem
29	523.4	27.7	525	22	AA515805 Human Six-Transmem
30	519	27.4	519	21	AAZ49398 Human STRAP-2 part
31	476.4	25.2	1213	21	AAZ46296 Human prostate gro
32	425.8	22.5	444	19	AAV61348 Forward DNA sequen
33	425.8	22.5	444	19	AAV58684 Prostate tumour sp
34	425.8	22.5	444	21	AAA06447 Human immunogenic
35	425.8	22.5	444	22	AA563655 Human prostate CDN
36	425.8	22.5	444	22	AA510206 Human prostate tum
37	425.8	22.5	444	22	AAH93563 Human prostate-spe
38	425.8	22.5	444	22	AAH84878 Human prostate-spe
39	425.8	22.5	444	22	AAH02628 Rat p-HYDE coding
40	406.6	21.5	2714	22	AAH89167 Ad5SVphye region
41	406.6	21.5	3885	22	AAH89169 Human foetal liver
42	395	20.9	395	22	ABA57377 Probe #5375 for ge
43	395	20.9	395	22	ABA26909 Human brain expres
44	395	20.9	395	22	AAK05412 Human bone marrow
45	395	20.9	395	22	AAK31011

## ALIGNMENTS

RESULT 1	
AA515801	standard; cDNA; 4329 BP.
ID	AA515801
AC	AA515801
XX	
DT	16-JAN-2002 (first entry)
DE	Human ORF of Six-Transmembrane Protein of Prostate 1, STMP1.
XX	
KW	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW	leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW	liver cancer; lung cancer; cytostatic; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	178..1650
FT	CDS
FT	/*tag= a
FT	/product= "STMP1"
XX	
PN	WO200172962-A2.
XX	
XX	04-OCT-2001.
PD	
XX	
XX	23-MAR-2001; 2001WO-US09410.
PE	
XX	
XX	24-MAR-2000; 2000US-191929P.
PR	
XX	
PA	(SAAT/) SAATCIGLU F.

XX Saactioglu F:  
PI WPI: 2001-662926/76.  
DR P-PSDB: AAU10187.  
XX  
PT New polynucleotide for the diagnosis, prevention and treatment for  
prostate and testis disorders, particularly prostate cancer, comprises  
prostate-specific or testis-specific nucleic acids  
XX  
PS Claim 5; Fig 4D; 114pp: English.  
XX  
CC The invention relates to substantially pure prostate-specific or  
testis-specific polypeptides and the nucleic acids encoding them.  
CC Also included are vectors and host cells expressing the proteins, a  
transgenic animal expressing the protein, antibodies against the  
CC proteins, probes for detecting the nucleic acids, antisense molecules  
CC for the nucleic acids and methods of isolating modulators of the  
CC proteins. Compounds that modulate the prostate specific or testis  
CC specific polypeptide are useful to diagnose, prevent or treat disorders  
CC of the testis or prostate particularly prostate cancer, benign  
CC prostatic hyperplasia, acute prostatitis, testicular cancer,  
CC cryptorchidism, undescended, retractile, ascending or vanished  
CC testis. Other proliferative disorders for which the modulators may be  
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast  
CC cancer, pancreatic cancer, liver cancer and lung cancer. The  
CC present sequence represents the open reading frame of a prostate  
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.  
XX  
SQ Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other:  
  
Query Match 85.1%; Score 1608.8; DB 22; Length 4329;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 675 tatatgcagcaacaatatattcaagcgcaacaaggtattatgaacttccgcagttgaa 734  
Oy 872 ttcatcccatgaacttggatcctctatcaacgcagagagattgaaattaccct 931  
Db 735 ttcatcccatgaacttggatcctctatcaacgcagagagattgaaattaccct 794  
Oy 932 acgacttctactctctgagagagcgagtggtgtagctataaagcttgcacatttt 991  
Db 795 acgacttctactctctgagagagcgagtggtgtagctataaagcttgcacatttt 854  
Oy 992 ttcccttattcctttgtcagaagatggtattcatccatatgtctagaaccacaagatga 1051  
Db 855 ttcccttattcctttgtcagaagatggtattcatccatatgtctagaaccacaagatga 914  
Oy 1052 ctcttacaanaattccctatagaagatttgaataaacccttacttattgttccatttt 1111  
Db 915 ctcttacaanaattccctatagaagatttgaataaacccttacttattgttccatttt 974  
Oy 1112 gctctccctagatataccctcgaaggtctctctgacgctgtctatcaacttattacggcac 1171  
Db 975 gctctccctagatataccctcgaaggtctctctgacgctgtctatcaacttattacggcac 1034  
Oy 1172 caagtatagagatttccaccttggcttggaaacctggtttacagtggtagaacaagcttgg 1231  
Db 1035 caagtatagagatttccaccttggcttggaaacctggtttacagtggtagaacaagcttgg 1094  
Oy 1232 attactaagtttttctctggtatggtccatggttgccttaccgctctgcttaccagatgg 1291  
Db 1095 attactaagtttttctctggtatggtccatggttgccttaccgctctgcttaccagatgg 1154  
Oy 1292 aaggttcagagagatatttcttctcaacatggtctatcagcaggtttcatgcanaatattga 1351  
Db 1155 aaggttcagagagatatttcttctcaacatggtctatcagcaggtttcatgcanaatattga 1214  
Oy 1352 aaactcttggaaatgaggaagatttggagaattggaattatatacccttggcataat 1411  
Db 1215 aaactcttggaaatgaggaagatttggagaattggaattatatacccttggcataat 1274  
Oy 1412 gaggcttggcttacttccctccctcggcagtcacttccctcagttgaggaattgttt 1471  
Db 1275 gaggcttggcttacttccctccctcggcagtcacttccctcagttgaggaattgttt 1334  
Oy 1472 aaacttggagagaattcagtttattcagctacacttgaatgctgctgtcctacaaag 1531  
Db 1335 aaacttggagagaattcagtttattcagctacacttgaatgctgctgtcctacaaag 1394  
Oy 1532 tacttccatggttttaattatgataatgaaacgagcttggaggaagatgactacagatt 1591  
Db 1395 tacttccatggttttaattatgataatgaaacgagcttggaggaagatgactacagatt 1454  
Oy 1592 ttataccccaacaaacttcttctgtctgttctgttgccttcaattgttaatttcgggttaa 1651  
Db 1455 ttataccccaacaaacttcttctgtctgttctgttgccttcaattgttaatttcgggttaa 1514  
Oy 1652 gatattttattccttccatgataaagcggaagcttaaaacgaatttaaaaaagcttggaa 1711  
Db 1515 gatattttattccttccatgataaagcggaagcttaaaacgaatttaaaaaagcttggaa 1574  
Oy 1712 aaagagcccaatttctggaagaaggtatattggaagaacaaatcccatgttcccccggagag 1771  
Db 1575 aaagagcccaatttctggaagaaggtatattggaagaacaaatcccatgttcccccggagag 1634  
Oy 1772 ggtcacagtaatgtagataaagtggttccacagctgcacataaagttctactatgc 1831  
Db 1635 ggtcacagtaatgtagataaagtggttccacagctgcacataaagttctactatgc 1694  
Oy 1832 catattattatgacttcaactcagttcagttacaaagtatgctgcanaattatcgttgggttga 1891  
Db 1695 catattattatgacttcaactcagttcagttacaaagtatgctgcanaattatcgttgggttga 1754

RESULT 2

ID	AA:15802	standard; cDNA; 1680 BP.
XX	AA:15802;	
XX	16-JAN-2002	(first entry)
XX		
DE	Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMPL.	
XX		
XX	Human; Six-Transmembrane Protein of Prostate 1; STMPL; prostate cancer;	
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;	
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;	
KW	leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;	
KW	liver cancer; lung cancer; cytostatic; ss.	
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	178..1650
FT	/*tag=	a
FT	/product=	"STMPL"
PN	MO200172962-A2.	
PD	04-OCT-2001.	
PE	23-MAR-2001; 2001WO-US09410.	
XX		
PR	24-MAR-2000; 2000US-191929P.	
XX		
PA	(SAAT/) SAATCLOGLU F.	
PI	Saactcloglu F;	
XX		
DR	WPI: 2001-662926/76.	
DR	P-PSDB; AAU10187.	
PT	New polynucleotide for the diagnosis, prevention and treatment for	
PT	prostate and testis disorders, particularly prostate cancer, comprises	
PT	prostate-specific or testis-specific nucleic acids	
XX		
PS	Claim 4; Fig 4E; 114pp; English.	
XX		
CC	The invention relates to substantially pure prostate-specific or	
CC	testis-specific polypeptides and the nucleic acids encoding them.	
CC	Also included are vectors and host cells expressing the proteins, a	
CC	transgenic animal expressing the protein, antibodies against the	
CC	proteins, probes for detecting the nucleic acids, antisense molecules	
CC	for the nucleic acids and methods of isolating modulators of the	
CC	proteins. Compounds that modulate the prostate specific or testis	
CC	specific polypeptide are useful to diagnose, prevent or treat disorders	
CC	of the testis or prostate particularly prostate cancer, benign	
CC	prostatic hyperplasia, acute prostatitis, testicular cancer,	
CC	cryptorchidism, undescended, retractile, ascending or vanished	
CC	testis. Other proliferative disorders for which the modulators may be	
CC	used include lymphoma, leukaemia, melanoma, ovarian cancer, breast	
CC	cancer, pancreatic cancer, liver cancer and lung cancer. The	
CC	present sequence encodes a prostate specific protein, Six-Transmembrane	
CC	Protein of Prostate 1, STMPL.	
XX		
SO	Sequence 1680 BP; 467 A; 334 C; 373 G; 506 T; 0 other;	
Query Match	81.2%	Score 1534.8; DB 22; Length 1680;
Best Local Similarity	99.5%	Pred. No. 0;
Matches 1539; Conservative	0;	Mismatches 7; Indels 0; Gaps 0;
272	atatcaagaagatattcttggtgatcttggaaagtgtccgtatcatggaatcaatctctat	331
11		
135	attgcgccagagatatctcttggtgatcttggaaagtgtccgtatcatggaatcaatctctat	194
332	gattgggaagcccttaagagccttagtgaacctgtttacccaatgacataatggtatcaaa	391

D	195	gaagggagagccccaagagccttgtagaacctgttacttaatgagataaagtgatca	254
Q	392	agatgcagaagaaagctcaactgtaagtgatctggaagtgagatcttgcgaactctgac	451
D	255	agatgcagaagaaagctcaactgtaagtgatctggaagtgagatcttgcgaactctgac	314
Q	452	cattcgacttattagatgagcggtcatactgtgtcattaggaagttagaattccaagtttgc	511
D	315	cattcgacttattagatgagcggtcatactgtgtcattaggaagttagaattccaagtttgc	374
Q	512	tttcgaattttttcccaatgtgtgatgtagtcatactcgtgaagaatgagctctcaaaaac	571
D	375	tctcgaaattcttccccaatgtgtgatgtagtcatactcgtgaagaatgagctctcaaaaac	434
Q	572	aaatataataattgtgtcctacacagagaaacctatactccctgtgagacctgagaca	631
D	435	aaatataataattgtgtcctacacagagaaacctatactccctgtgagacctgagaca	494
Q	632	tctgctgtggttaaaaatcctgtatgtagtgcgaataaacaatgagataaaccagtaacc	691
D	495	tctgctgtggttaaaaatcctgtatgtagtgcgaataaacaatgagataaaccagtaacc	554
Q	692	agaatccaatgctggaataatttgcttcattatctccagaattctcttgattgtcaaaagat	751
D	555	agaatccaatgctggaataatttgcttcattatctccagaattctcttgattgtcaaaagat	614
Q	752	taagtgtgtccagctctgagcgacctcaagttagagaccataagatgacgcgcgcaggtta	811
D	615	taagtgtgtccagctctgagcgacctcaagttagagaccataagatgacgcgcgcaggtta	674
Q	812	tatatgcagcaacaataattcaagcgcgcgaacaacaggttatgtgaacttgcgcgcagttga	871
D	675	tatatgcagcaacaataattcaagcgcgcgaacaacaggttatgtgaacttgcgcgcagttga	734
Q	872	tttatactccattggaacttgtagatccctatcatcaacgcagagaaattgaaatttaccct	931
D	735	tttatactccattggaacttgtagatccctatcatcaacgcagagaaattgaaatttaccct	794
Q	932	acgaactcttactctcgcggagagcgccagctgctggtgagtataagctcttgacacatttt	991
D	795	acgaactcttactctcgcggagagcgccagctgctggtgagtataagctcttgacacatttt	854
Q	992	tttccttattcctctgtgtagagatgtagtaatcatatgctaagaaccaaagatgta	1051
D	855	tttccttattcctctgtgtagagatgtagtaatcatatgctaagaaccaaagatgta	914
Q	1052	cttttaaaaaattcctataagaagtgtgtaataaaccttactatagtgtgcattacttt	1111
D	915	cttttaaaaaattcctataagaagtgtgtaataaaccttactatagtgtgcattacttt	974
Q	1112	gctctccctagtaacctgcgcaggtctctgcgcagctgtctatcaactcttatacgcgcac	1171
D	975	gctctccctagtaacctgcgcaggtctctgcgcagctgtctatcaactcttatacgcgcac	1034
Q	1172	caagatagagaaattccacactggtgtggaacactggttacagtgtagaanaacagcttgc	1231
D	1035	caagatagagaaattccacactggtgtggaacactggttacagtgtagaanaacagcttgc	1094
Q	1232	attactaagttttttctctgcctatgttccatagtgtgcttaacgcctctgtttacgcgatag	1291
D	1095	attactaagttttttctctgcctatgttccatagtgtgcttaacgcctctgtttacgcgatag	1154
Q	1292	aaggtcagagaaatattgtttctcaacaatgagcttaacgcaggttcatgcaaatatgta	1351
D	1155	aaggtcagagaaatattgtttctcaacaatgagcttaacgcaggttcatgcaaatatgta	1211
Q	1352	aaactcttggaatgaggaagaagtcttggaatctggaatgtaatactccttgcgataat	1411
D	1215	aaactcttggaatgaggaagaagtcttggaatctggaatgtaatactccttgcgataat	1274
Q	1412	gagccttggtctacttctcctcctgtgcagttcaacttatacccttcagtgagaaatgcttt	1471
D	1275	gagccttggtctacttctcctcctgtgcagttcaacttatacccttcagtgagaaatgcttt	1334

OY	1472	aaactgagagaatcaagtttatcaagtcacacttgatagtcgtctgtcataag	1531
Db	1335	aaactggagagaatctcagttttatcacgtacacttgatattgctgcctgcataag	1394
OY	1532	tacttccatggtttaattatgatgagaaacgagcttttgggaaggtactacaagt	1591
Db	1395	tacttccatggtttaattatgatgagaaacgagcttttgggaaggtactacaagt	1454
OY	1592	ttaatacacaacaactttctctgccttctgtttgcctcaattgtaattctggtaa	1651
Db	1455	ttatacaacaacaactttctctgccttctgtttgcctcaattgtaattctggtaa	1514
OY	1652	gattattttatctcctccatcgtataagccggaagcgtaaacgaattaaaaagcttggaa	1711
Db	1515	gattattttatctcctccatcgtataagccggaagcgtaaacgaattaaaaagcttggaa	1574
OY	1712	aaagaagccaattctctggaagaaggtatctggagaaacaattcctcatgtctcccggaag	1771
Db	1575	aaagaagccaattctctggaagaaggtatctggagaaacaattcctcatgtctcccggaag	1634
OY	1772	ggtacacgataatgatagaataatggtgttcaacgtgcgcataata	1817
Db	1635	ggtacacgataatgatagaataatggtgttcaacgtgcgcataata	1680

RESULT	3
AAS15793	
ID	AAS15793 standard; DNA; 1725 BP.
XX	
AC	AAS15793;
XX	
DT	16-JAN-2002 (first entry)
DE	
Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.	
XX	
KW	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW	kryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW	leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW	liver cancer; lung cancer; cyclostatic; ds.
XX	
OS	Homo sapiens.
FX	
Key	Location/Qualifiers
FT	89
FT	/tag= a
FT	/note= "Represents 338 nucleotides of intron 1"
FT	162
FT-	/tag= b
FT	/note= "Represents 12713 nucleotides of intron 2"
FT	200..1702
FT	/tag= c
FT-	/product= "STMP1"
FT	697
FT	/tag= d
FT	/note= "Represents 1396 nucleotides of intron 3"
FT	1225
FT	/tag= e
FT	/note= "Represents 2372 nucleotides of intron 4"
FT	1410
FT	/tag= f
FT	/note= "Represents 2299 nucleotides of intron 5"
PN	
MO200172962-A2.	
XX	
PD	04-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09410.
XX	
PR	24-MAR-2000; 2000US-191929P.
XX	
PA	(SAAT/) SAATCIGLU F.

XX Saactcloglu F;  
PI  
XX  
XX  
DR WPI: 2001-662926/76.  
DR P-PSDB: ANU10187.  
XX  
XX New polynucleotide for the diagnosis, prevention and treatment for  
PT prostate and testis disorders, particularly prostate cancer, comprises  
PT prostate-specific or testis-specific nucleic acids  
PT  
XX  
XX Example 3; Fig 4B; 11App; English.  
XX  
XX The invention relates to substantially pure prostate-specific or  
CC testis-specific polypeptides and the nucleic acids encoding them.  
CC Also included are vectors and host cells expressing the proteins, a  
CC transgenic animal expressing the protein, antibodies against the  
CC proteins, probes for detecting the nucleic acids, antisense molecules  
CC for the nucleic acids and methods of isolating modulators of the  
CC proteins. Compounds that modulate the prostate specific or testis  
CC specific polypeptide are useful to diagnose, prevent or treat disorders  
CC of the testis or prostate particularly prostate cancer, benign  
CC prostatic hyperplasia, acute prostatitis, testicular cancer,  
CC cryptorchidism, undescended, retractile, ascending or vanished  
CC testis. Other proliferative disorders for which the modulators may be  
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast  
CC cancer, pancreatic cancer, liver cancer and lung cancer. The  
CC present sequence encodes a prostate specific protein, Six-Transmembrane  
CC Protein of Prostate 1, STMP1.  
XX  
XX Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other:  
XX

Query Match	77.5%	Score 1466.2	DB 22	Length 1725
Best Local Similarity	97.9%	Pred. No. 0		
Matches 1528	Conservative	0	Mismatches 3	Indels 30
				Gaps 3
QY 280	agagatactctgcatcttggaatggtcccgatcatcagatcaatcctatgatgaggaa	339		
DB 165	agagatactctgcatcttggaatggtcccgatcatcagatcaatcctatgatgaggaa	224		
QY 340	ggccctaaagagcctagtggaactggttaacctatgagcaaaatggtatcaaaagtgcga	399		
DB 225	ggccctaaagagcctagtggaactggttaacctatgagcaaaatggtatcaaaagtgcga	284		
QY 400	ggaaaggtccacttaagtgtgatttggaagtgaggagatttgccaaatccctggacattcgac	459		
DB 285	ggaaaggtccacttaagtgtgatttggaagtgaggagatttgccaaatccctggacattcgac	344		
QY 460	ttatagatgcgcgtacatcgtggtcatatgagaagtgaatccctaaagtttgtcttcgaat	519		
DB 345	ttatagatgcgcgtacatcgtggtcatatgagaagtgaatccctaaagtttgtcttcgaat	404		
QY 520	tttttcctcatgvtggttagatgctcaactcatcatgaaagtctctcacaaaaacaataata	579		
DB 405	tttttcctcatgvtggttagatgctcaactcatcatgaaagtctctcacaaaaacaataata	464		
QY 580	tatttgtgctataacagagaagaacattactccctctgvggagaccgtgaaactctgcttg	639		
DB 465	tatttgtgctataacagagaagaacattactccctctgvggagaccgtgaaactctgcttg	524		
QY 640	tgggtcaaaatccctgatttgaatgtgagacaataacatgtaggtcaaacccagatccca	699		
DB 525	tgggtcaaaatccctgatttgaatgtgagacaataacatgtaggtcaaacccagatccca	584		
QY 700	atgcgtcaaatatttgcttcatatattcccaagattcttggattgltcacaagagattaaatg	759		
DB 585	atgcgtcaaatatttgcttcatatattcccaagattcttggattgltcacaagagattaaatg	644		
QY 760	tctcagcttgggcaactcagttagaacctaaagatgcccagccgg-----caggtt	809		
DB 645	tctcagcttgggcaactcagttagaacctaaagatgcccagccaggatgataagacaggtt	704		
QY 810	tatatatgaagacaacatatcaacgcgcgcacacagatcatgtaactgtccgcacagttg	869		

```

Db 705 tataatagcagcaaatattcaagcgcgacacaggttaattgacttgcgcgcaggttg 764
      |||
Qy 870 aatttattccattgacttggatggtatccattatcaccgagagagattgaaattacc 929
      |||
Db 765 aatttattccattgacttggatggtatccattatcaccgagagagattgaaattacc 824
      |||
Qy 930 ctacgactcttactctctcggagagagcagtggtgtaactaagcttgccacattt 989
      |||
Db 825 ctacgactcttactctctcggagagagcagtggtgtaactaagcttgccacattt 884
      |||
Qy 990 ttttcttattctctcttgcagagatgtgattcattccatgtcagaaaccacagagt 1049
      |||
Db 885 ttttcttattctctcttgcagagatgtgattcattccatgtcagaaaccacagagt 944
      |||
Qy 1050 gacttttaacaaattcctatagagattgtgataaaaccttactattagtgtccattac 1109
      |||
Db 945 gacttttaacaaattcctatagagattgtgataaaaccttactattagtgtccattac 1004
      |||
Qy 1110 ttgctctccctagatatacctcgcaggtctcttgcagcgtgctataacttattacg 1169
      |||
Db 1005 ttgctctccctagatatacctcgcaggtctcttgcagcgtgctataacttattacg 1064
      |||
Qy 1170 accaaglatagagagattccacacttgttggaaacctgtgtacagtgtagaanaacagctt 1229
      |||
Db 1065 accaaglatagagagattccacacttgttggaaacctgtgtacagtgtagaanaacagctt 1124
      |||
Qy 1230 ggaattcctaagtttctctcgtcattgctcattgctcagcctctgctaccagatg 1289
      |||
Db 1125 ggaattcctaagtttctctcgtcattgctcattgctcagcctctgctaccagatg 1184
      |||
Qy 1290 agaaggtcagaagagatattgtttcacaatgcttaccagc-----aggttca 1339
      |||
Db 1185 agaaggtcagaagagatattgtttcacaatgcttaccagc-----aggttca 1244
      |||
Qy 1340 tgaagaatattgaaacctcttggaatagagaagagatttggagaatttgaatgtatctc 1399
      |||
Db 1245 tgaagaatattgaaacctcttggaatagagaagagatttggagaatttgaatgtatctc 1304
      |||
Qy 1400 ctttggcataatgagccttggcttacttccctcctgcagcacttctatccctcagt 1459
      |||
Db 1305 ctttggcataatgagccttggcttacttccctcctgcagcacttctatccctcagt 1364
      |||
Qy 1460 gagcaatgctttaacttgagagaattcagtttattcag-----tcaaccttg 1509
      |||
Db 1365 gagcaatgctttaacttgagagaattcagtttattcag-----tcaaccttg 1424
      |||
Qy 1510 gataatgcctcgcctcacaataagtaacttccatgtttaaattatgataggaacagagctt 1569
      |||
Db 1425 gataatgcctcgcctcacaataagtaacttccatgtttaaattatgataggaacagagctt 1484
      |||
Qy 1570 ttgaagaagaagtaactaagatttataacacacaaaccttgttcttgcctgttttgc 1629
      |||
Db 1485 ttgaagaagaagtaactaagatttataacacacaaaccttgttcttgcctgttttgc 1544
      |||
Qy 1630 cctcaatttgaattcctgggtaagattatttattccttccatgtataagccggaactaa 1689
      |||
Db 1545 cctcaatttgaattcctgggtaagattatttattccttccatgtataagccggaactaa 1604
      |||
Qy 1690 aacgaatttaaaaggcctggaaaagagcaatttcttggaagaagttatggggagacaa 1749
      |||
Db 1605 aacgaatttaaaaggcctggaaaagagcaatttcttggaagaagttatggggagacaa 1664
      |||
Qy 1750 ttccctcatgtctcccgagagaggttcacagtaattgtatgataaagtgtgtcacagctg 1809
      |||
Db 1665 ttccctcatgtctcccgagagaggttcacagtaattgtatgataaagtgtgtcacagctg 1724
      |||
Qy 1810 c 1810
      |||
Db 1725 c 1725
      |||

```

RESULT 4

```

AAD07072
ID AAD07072 standard; cDNA; 2453 BP.
XX AC AAD07072;
XX AC 06-AUG-2001 (first entry)
XX DE Human six transmembrane epithelial antigen of prostate-2 clone GND3 cDNA.
XX XX Human; cytosolic; antiproliferative; vaccine; gene therapy;
XX KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;
XX KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
XX KW pancreatic; ss.
XX OS Homo sapiens.
XX EH Key Location/Qualifiers
XX EH RBS 352..358
XX FT /tag= a
XX FT /note= "Kozak region"
XX FT CDS 355..1719
XX FT /tag= b
XX FT /product= "Human six transmembrane epithelial antigen
XX FT of the prostate (STEAP)-2"
XX FT CDS 709..2073
XX FT /tag= c
XX FT /product= "Human six transmembrane epithelial antigen
XX FT of the prostate (STEAP)-2, alternative version"
XX FT /note= "CDS does not include start and stop codon"
XX FT /transl_except= (pos:1714..1722, aa:Asp-Ala)
XX FT /transl_except= (pos:1834..1842, aa:Arg-Ser)
XX FT /transl_except= (pos:1957..1965, aa:Glu-Gly)
XX FT /transl_except= (pos:2050..2058, aa:Thr-Ser)
XX FT /transl_except= (pos:2062..2070, aa:Asn-Phe)
XX FT /note= "inframe stop codon alters the reading frame"
XX FT /partial
XX PN WO200140276-A2.
XX PD 07-JUN-2001.
XX PF 06-DEC-2000; 2000WO-US33040.
XX PR 06-DEC-1999; 99US-0455486.
XX PA (UROC-) UROGENESYS INC.
XX PI Afar DEH, Hubert RS, Raitano AB, Safran DC, Mitchell SC, Paris M;
XX PI Jakobovits A;
XX XX WPI; 2001-367804/38.
XX DR P-PSDB; AAE02781, AAE02841.
XX PT New STEAP (six transmembrane epithelial antigen of the prostate)
XX PT proteins, expressed in human cancers, useful for detecting and treating
XX PT cancer.
XX PS Claim 4; Fig 9A-9D; 187pp; English.
XX XX
XX CC The present sequence is human six transmembrane epithelial antigen of
XX CC the prostate (STEAP)-2 clone GND3 cDNA. STEAP is a member of cell
XX CC surface serpentine transmembrane antigens. STEAP-2 gene is located on
XX CC chromosome 7q21 and is used in gene therapy. Inhibiting the development
XX CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
XX CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
XX CC expressing STEAP in a patient, comprises administering a vaccine
XX CC composition to the patient. Treating a patient with a cancer that
XX CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
XX CC comprises administering to the patient a vector encoding single chain
XX CC monoclonal antibody that comprises the variable domains of the heavy and
XX CC light chains of the monoclonal antibody that specifically binds to STEAP,
XX CC such that the vector delivers the single chain monoclonal antibody coding
XX CC sequence to the cancer cells and the encoded single chain monoclonal

```



XX MO200172962-A2.  
XX 04-OCT-2001.  
XX  
XX 23-MAR-2001; 2001MO-US09410.  
XX  
XX 24-MAR-2000; 2000US-191929P.  
XX  
XX (SAAT/) SAATC10GLU F.  
XX  
XX Saatl0glu F;  
XX  
XX MPI; 2001-662926/76.  
XX P-PSDB; AAU10188.  
XX  
XX New polynucleotide for the diagnosis, prevention and treatment for  
XX prostate and testis disorders, particularly prostate cancer, comprises  
XX prostate-specific or testis-specific nucleic acids  
XX  
XX Claim 4; Fig 4H; 11App; English.  
XX  
XX The invention relates to substantially pure prostate-specific or  
XX testis-specific polypeptides and the nucleic acids encoding them.  
XX Also included are vectors and host cells expressing the proteins, a  
XX transgenic animal expressing the protein, antibodies against the  
XX proteins, probes for detecting the nucleic acids, antisense molecules  
XX for the nucleic acids and methods of isolating modulators of the  
XX proteins. Compounds that modulate the prostate specific or testis  
XX specific polypeptide are useful to diagnose, prevent or treat disorders  
XX of the testis or prostate particularly prostate cancer, benign  
XX prostatic hyperplasia, acute prostatitis, testicular cancer,  
XX cryptorchidism, undescended, retractile, ascending or vanished  
XX testis. Other proliferative disorders for which the modulators may be  
XX used include lymphoma, leukemia, melanoma, ovarian cancer, breast  
XX cancer, pancreatic cancer, liver cancer and lung cancer. The  
XX present sequence encodes prostate specific protein, Six-Transmembrane  
XX protein of Prostate 1, STMP1, ORF2.  
XX  
XX Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;  
SQ

Query Match 72.4%; Score 1369.2; DB 22; Length 1561;  
Best Local Similarity 99.8%; Pred. 0;  
Matches 1371; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 276 tcaagagatattcttgtagcttggaaagtgctcgatcatggaatcaatctctatgatg 335  
DB 149 tccagagatatctctgtgcttggaagtgctcgatcatggaatcaatctctatgatg 208

QY 336 ggaagccctaagagccttagtgaaactgtttaccctaagtgcataatggtatcaaat 395  
DB 209 ggaagccctaagagccttagtgaaactgtttaccctaagtgcataatggtatcaaat 268

QY 396 gcaaggaagtcacgttagtgatgtgaaagtggagatttgcgaatccttgaccatt 455  
DB 269 gcaaggaagtcacgttagtgatgtgaaagtggagatttgcgaatccttgaccatt 328

QY 456 cgaactattagatgcgcatcatcatgtgcataggaagtagaactctaagtgtcttct 515  
DB 329 cgaactattagatgcgcatcatcatgtgcataggaagtagaactctaagtgtcttct 388

QY 516 gaatttttctcatgtgtaagatgtcaccatcatgagaagatgctctcaaaaaaacaat 575  
DB 389 gaatttttctcatgtgtaagatgtcaccatcatgagaagatgctctcaaaaaaacaat 448

QY 576 ataataattgtctatcacagaacatataactccctcggggagcctgagacatctg 635  
DB 449 ataataattgtctatcacagaacatataactccctcggggagcctgagacatctg 508

QY 636 ctgtggtgtaaaactcctgattgattgagcaataacatgagataaaccagtaaccagaa 695  
DB 509 ctgtggtgtaaaactcctgattgattgagcaataacatgagataaaccagtaaccagaa 568

QY 696 tccaatgctgaatatattggtctcatattccagatcttggattgtgcaaaagatttaac 755  
DB 569 tccaatgctgaatatattggtctcatattccagatcttggattgtgcaaaagatttaac 628

QY 756 gttgtcctcagcttgggcaacttcagtaggaacctaaagatgcagcgaggtttatata 815  
DB 629 gttgtcctcagcttgggcaacttcagtaggaacctaaagatgcagcgaggtttatata 688

QY 816 tgcagcaacaatatattcaagcgagcaacaagattatgaatttgcgcgaggtttgaattc 875  
DB 689 tgcagcaacaatatattcaagcgagcaacaagattatgaatttgcgcgaggtttgaattc 748

QY 876 attcccatgacttggagatccttatcatcagcaggaagattgaaatttaacctcaga 935  
DB 749 attcccatgacttggagatccttatcatcagcaggaagattgaaatttaacctcaga 808

QY 936 ctcttactctctggaagagggccagtggtgtagtataagcttggccacattttttc 995  
DB 809 ctcttactctctggaagagggccagtggtgtagtataagcttggccacattttttc 868

QY 996 cttaattccttgcagagatggtatcatcatatgctagaaaccaagagtgacttt 1055  
DB 869 cttaattccttgcagagatggtatcatcatatgctagaaaccaagagtgacttt 928

QY 1056 tacaataattcctatagagattggaataaaacttaccattagtggcatttctgctc 1115  
DB 929 tacaataattcctatagagattggaataaaacttaccattagtggcatttctgctc 988

QY 1116 tccctagtaactctgcagagcttctctgcagcgtctataacttatlacgagcaaacg 1175  
DB 989 tccctagtaactctgcagagcttctctgcagcgtctataacttatlacgagcaaacg 1048

QY 1176 tatagaagattccacctgtgttggaacctggtttacagtgtagaagaaacagcttgatta 1235  
DB 1049 tatagaagattccacctgtgttggaacctggtttacagtgtagaagaaacagcttgatta 1108

QY 1236 cttaagtttttctctgctagatgctcagtggtgtgtctacagcctctgcttaccagtagaagg 1295  
DB 1109 cttaagtttttctctgctagatgctcagtggtgtgtgtgtctacagcctctgcttaccagtagaagg 1168

QY 1296 tcagagagatatattgtttctcaacatggtctatcagcaggttcatggaatatgaaac 1355  
DB 1169 tcagagagatatattgtttctcaacatggtctatcagcaggttcatggaatatgaaac 1228

QY 1356 tcttggaaatggaagaaggttggagaattgaaatgtatatactccttggcataatgagc 1415  
DB 1229 tcttggaaatggaagaaggttggagaattgaaatgtatatactccttggcataatgagc 1288

QY 1416 ctggcttacttccctcccgagcagatccttatacccttcaatggagcaatgctttaaac 1475  
DB 1289 ctggcttacttccctcccgagcagatccttatacccttcaatggagcaatgctttaaac 1348

QY 1476 tggagagaattcagtttatttcagctcacacttggaatagtctcctgcaccataagtaact 1535  
DB 1349 tggagagaattcagtttatttcagctcacacttggaatagtctcctgcaccataagtaact 1408

QY 1536 ttccatgttttaattatggaatggaacagagcttttggagaagagtaactacaattttat 1595  
DB 1409 ttccatgttttaattatggaatggaacagagcttttggagaagagtaactacaattttat 1468

QY 1596 acacacacaaacttttctgtcttggtttggtttgcctcccaattttaaatttcagggt 1649  
DB 1469 acacacacaaacttttctgtcttggtttggtttgcctcccaattttaaatttcagggt 1522

RESULT 6  
AA515810  
ID AA515810 standard; cDNA: 2238 BP.  
XX  
AC AA515810;  
XX  
DT 16-JAN-2002 (first entry)



XX	Human ORF2 of Six-Transmembrane Protein of Prostate 1, STMP1.
DE	
KW	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW	leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW	liver cancer; lung cancer; cytosolic; ss; ORF2.
XX	
OS	Homo sapiens.
XX	
FI	Key
FI	Location/Qualifiers
FI	CDS
FI	188..1552
FI	/tag= a
FI	/product= "STMP1, ORF2"
XX	
PN	WO200172962-A2.
XX	
PD	04-OCT-2001.
XX	
PE	23-MAR-2001; 2001WO-US09410.
XX	
PR	24-MAR-2000; 2000US-191929P.
XX	
PA	(SAAT/) SAATC10GLU F.
XX	
PI	Saatc10glu F;
XX	
DR	WPI: 2001-662926/76.
DR	P-PSDB; AA010188.
XX	
PT	New polynucleotide for the diagnosis, prevention and treatment for
PT	prostate and testis disorders, particularly prostate cancer, comprises
PT	prostate-specific or testis-specific nucleic acids
XX	
PS	Claim 5; Fig 4G; 114pp; English.
XX	
CC	The invention relates to substantially pure prostate-specific or
CC	testis-specific polypeptides and the nucleic acids encoding them.
CC	Also included are vectors and host cells expressing the proteins, a
CC	transgenic animal expressing the protein, antibodies against the
CC	proteins, probes for detecting the nucleic acids, antisense molecules
CC	for the nucleic acids and methods of isolating modulators of the
CC	proteins. Compounds that modulate the prostate specific or testis
CC	specific polypeptide are useful to diagnose, prevent or treat disorders
CC	of the testis or prostate particularly prostate cancer, benign
CC	prostatic hyperplasia, acute prostatitis, testicular cancer,
CC	cryptorchidism, undescended, retractile, ascending or vanished
CC	testis. Other proliferative disorders for which the modulators may be
CC	used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC	cancer, pancreatic cancer, liver cancer and lung cancer. The
CC	present sequence represents the second open reading frame of a prostate
CC	specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX	
Q0	Sequence 2238 BP; 607 A; 457 C; 453 G; 721 T; 0 other;

Query Match	72.3%	Score 1367.6	DB 22	Length 2238
Best Local Similarity	99.7%	Pred. No. 0		
Matches 1370	Conservative	0	Mismatches 4	Indels 0
				Gaps 0
QY 276	tcaagatattcttgggacatcttggaaagtctgcgtatcatcggaaatcaatctctatgatg	335		
Db 149	tcccaagatattcttgggacatcttggaaagtctgcgtatcatcggaaatcaatctctatgatg	208		
QY 336	ggaagccctaagagccttaagtgaaactggtttacctaatgtgcataaagatgatacaaat	395		
Db 209	ggaagccctaagagccttaagtgaaactggtttacctaatgtgcataaagatgatacaaat	268		
QY 396	gcaagaaaggtcacatgtatgtgtatgttggaaagtggagaatttggccaatctccttgaccatt	455		
Db 269	gcaagaaaggtcacatgtatgtgtatgttggaaagtggagaatttggccaatctccttgaccatt	328		

QY	456	cgactatataatgcgctatcatgctgataatggaagttagaaatcccaagttgcttc	515
Dd	329	cgactatataatgcgctatcatgctgataatggaagttagaaatcccaagttgcttc	388
QY	516	gaattttccctaatgtagatgataatgcatacaataaataatctcccaaaaacaaat	575
Dd	389	gaattttccctaatgtagatgataatgcatacaataaataatctcccaaaaacaaat	448
QY	576	ataaatgttgtcatalatacaagaacattataacccctcgttggagcttgacaatctg	635
Dd	449	ataaatgttgtcatalatacaagaacattataacccctcgttggagcttgacaatctg	508
QY	636	cttggggtataaaccctgatgtgatgtagaataaataatgagataaaccagatcccaaga	695
Dd	509	cttggggtataaaccctgatgtgatgtagaataaataatgagataaaccagatcccaaga	568
QY	696	ccaatgttgataatttgagctcaatattcccaagatcttgtattgttcaagaatttaac	755
Dd	569	ccaatgttgataatttgagctcaatattcccaagatcttgtattgttcaagaatttaac	628
QY	756	gttgctcgaagcttgggcacactcagtttagaacctaagaatgccaagccggcaggttatata	815
Dd	629	gttgctcgaagcttgggcacactcagtttagaacctaagaatgccaagccggcaggttatata	688
QY	816	tgcagcaacaatattcaagcgcgaacaacagtgattatgaacttgcgcgcagattgaattc	875
Dd	689	tgcagcaacaatattcaagcgcgaacaacagtgattatgaacttgcgcgcagattgaattc	748
QY	876	attcccaattgagcttggagatccctataccaagccagaagagatgtaaaaattaccctaaga	935
Dd	749	attcccaattgagcttggagatccctataccaagccagaagagatgtaaaaattaccctaaga	808
QY	936	ctcttactctcttggagaaggccagttgtgtatgataaagcttggccaattttcttc	995
Dd	809	ctcttactctcttggagaaggccagttgtgtatgataaagcttggccaattttcttc	868
QY	996	ctttattcccttggccaagaatgtgatccaatcacaatgctagaaaccaagaagtgaactt	1055
Dd	869	ctttattcccttggccaagaatgtgatccaatcacaatgctagaaaccaagaagtgaactt	928
QY	1056	tacaaaattccctatagaagatgtgtgaataaanocttaccatagttgccaattctgtgc	1115
Dd	929	tacaaaattccctatagaagatgtgtgaataaanocttaccatagttgccaattctgtgc	988
QY	1116	tccctaglataaccttcgaagctctctcttgcgcagctgcattacaacttattatagccaccaag	1175
Dd	989	tccctaglataaccttcgaagctctctcttgcgcagctgcattacaacttattatagccaccaag	1048
QY	1176	tatagsgaattccaacttgtgttggaacacttgttaccagttagaanaacagcttggatta	1235
Dd	1049	tatagsgaattccaacttgtgttggaacacttgttaccagttagaanaacagcttggatta	1108
QY	1236	ctaagttttttcttcgataatgataatgataatgataatgataatgataatgataatgataatg	1295
Dd	1109	ctaagttttttcttcgataatgataatgataatgataatgataatgataatgataatgataatg	1168
QY	1296	tcaagaagataattgtttcccaacacagctgcattacaagcaggttccatgcataatgtgaanaac	1355
Dd	1169	tcaagaagataattgtttcccaacacagctgcattacaagcaggttccatgcataatgtgaanaac	1228
QY	1356	tctctgtaagttagaagaagtttggagaatgtaaatgtatatctccttggcataaagtgc	1415
Dd	1229	tctctgtaagttagaagaagtttggagaatgtaaatgtatatctccttggcataaagtgc	1288
QY	1416	cttggcttacttccctcccggaagccaacttcatctatcccttcagtgagcaatgctttaaac	1475
Dd	1289	cttggcttacttccctcccggaagccaacttcatctatcccttcagtgagcaatgctttaaac	1348
QY	1476	tggagagaatttcaagtttattatcagttacaacttgataatgtctgccttcataaagtact	1535
Dd	1349	tggagagaatttcaagtttattatcagttacaacttgataatgtctgccttcataaagtact	1408
QY	1536	tccaagtctttaaatttatagatgtagaaacagacttcttgagagaagataacacagatttat	1595



DB 1409 ttccatgtttaattatgatgagaaagacgttttgaggaagagactacagatttat 1468  
QY 1596 aacaccacaaacttgcttctgtctgtttgcccataatgtatcttgat 1649  
DB 1469 aacaccacaaacttgcttctgtctgtttgcccataatgtatcttgat 1522

## RESULT 7

AA515813  
ID AA515813 standard; cDNA; 2102 BP.

AA515813;

16-JAN-2002 (first entry)

Human cDNA encoding ORF3 of Six-Transmembrane Protein of Prostate 1.

Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;  
benign prostatic hyperplasia; acute prostatitis; testicular cancer;  
cryptorchidism; testicular disorder; proliferative disorder; lymphoma;  
leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;  
liver cancer; lung cancer; cytostatic; SS; ORF3.

Homo sapiens.

Key Location/Qualifiers

FT CDS 171..1430

FT /tag= a

FT /product= "STMP1, ORF3"

WO200172962-A2.

04-OCT-2001.

23-MAR-2001; 2001WO-US09410.

24-MAR-2000; 2000US-191929P.

(SAAT/) SAATCTOGIU F.

Saatctoglu F;

WPI: 2001-662926/76.

P-PSDB: AANU10189.

New polynucleotide for the diagnosis, prevention and treatment for prostate and testis disorders, particularly prostate cancer, comprises prostate-specific or testis-specific nucleic acids

Claim 4; Fig 4K; 114pp; English.

The invention relates to substantially pure prostate-specific or testis-specific polypeptides and the nucleic acids encoding them. Also included are vectors and host cells expressing the proteins, a transgenic animal expressing the protein, antibodies against the proteins, probes for detecting the nucleic acids, antisense molecules for the nucleic acids and methods of isolating modulators of the proteins. Compounds that modulate the prostate specific or testis specific polypeptide are useful to diagnose, prevent or treat disorders of the testis or prostate particularly prostate cancer, benign prostatic hyperplasia, acute prostatitis, testicular cancer, CC cryorchidism, undescended, retractile, ascending or vanished testis. Other proliferative disorders for which the modulators may be used include lymphoma, leukaemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, liver cancer and lung cancer. The present sequence encodes prostate specific protein, Six-Transmembrane Protein of Prostate 1, STMP1, ORF3.

Sequence 2102 BP; 595 A; 424 C; 422 G; 661 T; 0 other;

Query Match 64.3%; Score 1216.8; DB 22; Length 2102;

Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1221; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 272 ataattcaagagatattcttggatctcttggaagtgctcgatatcatatgaaatctctat 331  
DB 128 atgtgcagagatattcttggatctcttggaagtgctcgatatcatatgaaatctctat 187  
QY 332 gatgggaagcccaagagccttagtgaactgtttacctaagcgataatggtatcaa 391  
DB 188 gatgggaagcccaagagccttagtgaactgtttacctaagcgataatggtatcaa 247  
QY 392 agatgcaaggaagtgctacgttagtgaactgttgaagtggaagatttgccaatccctgac 451  
DB 248 agatgcaaggaagtgctacgttagtgaactgttgaagtggaagatttgccaatccctgac 307  
QY 452 catcgactattagatgctggctatacatgtgtgcataggaagtgaataatcctaagttgc 511  
DB 308 catcgactattagatgctggctatacatgtgtgcataggaagtgaataatcctaagttgc 367  
QY 512 ttctgaatttttcccatggtgtagatgtacatacctaagatgctctcccaaaac 571  
DB 368 ttctgaatttttcccatggtgtagatgtacatacctaagatgctctcccaaaac 427  
QY 572 aaataataatgtgtgtacatacacagagaacacataaacctccgttggaacctgagaa 631  
DB 428 aaataataatgtgtgtacatacacagagaacacataaacctccgttggaacctgagaa 487  
QY 632 tctgtctgtgggtaaatccctgatgtatgtgagcaataacatgagataaccagttacc 691  
DB 488 tctgtctgtgggtaaatccctgatgtatgtgagcaataacatgagataaccagttacc 547  
QY 692 agaattccaatgtcgaatattgctcattatcccaattcttgattgtccaagatt 751  
DB 548 agaattccaatgtcgaatattgctcattatcccaattcttgattgtccaagatt 607  
QY 752 taatgtgtcctcagcttggcgaacttcaagttagaacctaaagatgccaagcgaagtta 811  
DB 608 taatgtgtcctcagcttggcgaacttcaagttagaacctaaagatgccaagcgaagtta 667  
QY 812 tatatgagacaacaataatccaagcgcgaacaacaggttatatgaaattgcccgaattgaa 871  
DB 668 tatatgagacaacaataatccaagcgcgaacaacaggttatatgaaattgcccgaattgaa 727  
QY 872 ttcatctccattgacttggatggtccttatcatcgaagcgaagatgaaatttaccct 931  
DB 728 ttcatctccattgacttggatggtccttatcatcgaagcgaagatgaaatttaccct 787  
QY 932 acgaacttactctcttgagagagggcgaagtgtgtgagctataagcttggccaatttt 991  
DB 788 acgaacttactctcttgagagagggcgaagtgtgtgagctataagcttggccaatttt 847  
QY 992 ttctcttatctccttgcgaagatgtgatatccatatgtcgaagaacaaagaagtga 1051  
DB 848 ttctcttatctccttgcgaagatgtgatatccatatgtcgaagaacaaagaagtga 907  
QY 1052 ctttacaataatcccatagagatgtgataaaccattaccatagttggccaatttt 1111  
DB 908 ctttacaataatcccatagagatgtgataaaccattaccatagttggccaatttt 967  
QY 1112 gctctccatgatatactcgcaggtctcttgcgaagtgttalcacatttatcagcgaac 1171  
DB 968 gctctccatgatatactcgcaggtctcttgcgaagtgttalcacatttatcagcgaac 1027  
QY 1172 caagtatagagatttccacacttggttgaacactgtgtacagtgtaagaanaacagcttg 1231  
DB 1028 caagtatagagatttccacacttggttgaacactgtgtacagtgtaagaanaacagcttg 1087  
QY 1232 attactaaagttttcttgcatagtgtcatgttgctcaacgactgtgttaccagttgag 1291  
DB 1088 attactaaagttttcttgcatagtgtcatgttgctcaacgactgtgttaccagttgag 1147  
QY 1292 aaggtcagaagatatgtttctcacaatggttatacgaaggtcatalgcaaatatgtga 1351



QY 961 tgggtgagctataagcttggccacatlttttcccttattcccttggtagagatgtga 1020  
|||||  
Db 941 tgggtgagctataagcttggccacatlttttcccttattcccttggtagagatgtga 1000  
QY 1021 ttcattcatatgtctgaagaacacagagtgactttcaaaaattccatagagatgtga 1080  
|||||  
Db 1001 ttcattcatatgtctgaagaacacagagtgactttcaaaaattccatagagatgtga 1060  
QY 1081 ataaacctactatagttgcttacttcttctctccctagatataactctgcaggtcttc 1140  
|||||  
Db 1061 ataaacctactatagttgcttacttcttctctccctagatataactctgcaggtcttc 1120  
QY 1141 tggcagctgcttatacaacttattatagcgcacacagatagagatccaccctgtgttg 1200  
|||||  
Db 1121 tggcagctgcttatacaacttattatagcgcacacagatagagatccaccctgtgttg 1180  
QY 1201 aaacctggttacaagtgtagaagaacagcttgatctactaaagttttcttcgtatgttc 1260  
|||||  
Db 1181 aaacctggttacaagtgtagaagaacagcttgatctactaaagttttcttcgtatgttc 1240  
QY 1261 atgttgcttacaagcttctgttccagtgtagaagtgtagagatattgtttctcaaca 1320  
|||||  
Db 1241 atgttgcttacaagcttctgttccagtgtagaagtgtagagatattgtttctcaaca 1300  
QY 1321 tggcttatacagcag 1334  
|||||  
Db 1301 tggcttatacagcag 1314

RESULT 9  
AAS64300  
ID AAS64300 standard; cDNA; 3900 BP.  
XX  
AC AAS64300:  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #104.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG00113.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 1; SEQ ID NO 104; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WPI  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3900 BP; 1161 A; 798 C; 892 G; 1049 T; 0 other;

Query Match 57.8%; Score 1093.2; DB 23; Length 3900;  
Best Local Similarity 91.3%; Pred. NO. 4.1e-314;  
Matches 1218; Conservative 0; Mismatches 3; Indels 113; Gaps 1;

QY 1 ggggaagcagctgtagtgcagccgacagccagccctgcacacccgacgtcgaggtg 60  
|||||  
Db 94 ggggaagcagctgtagtgcagccgacagccagccctgcacacccgacgtcgaggtg 153  
QY 61 caatcgtagccctgagcccgagggtggcccttgggaagtgcggccgctcccgaggag 120  
|||||  
Db 154 caatcgtagccctgagcccgagggtggcccttgggaagtgcggccgctcccgaggag 213  
QY 121 ctgcaagctcgcccgcccgccgctgtagagcgcgaggggcgcgagggaagtgaagaga 180  
|||||  
Db 214 ctgcaagctcgcccgcccgccgctgtagagcgcgaggggcgcgagggaagtgaagaga 259  
QY 181 ggaattggaattgtgagtgaccttcgtatcgtctcctctgctggaagaagggg 240  
|||||  
Db 260 ----- 259  
QY 241 aagaactgcatgcatattatcagcgtcctataatccaagagatattctgtgacttg 300  
|||||  
Db 260 -----agatatcttctgtgacttg 280  
QY 301 gaagtgcgtatcatatggaatcattctatgtaggaagccctaaagagcttagtga 360  
|||||  
Db 281 gaagtgcgtatcatatggaatcattctatgtaggaagccctaaagagcttagtga 340  
QY 361 ctgtttaccatagtgataatgtagtcaagaatgcaaggaaggtactgtagtga 420  
|||||  
Db 341 ctgtttaccatagtgataatgtagtcaagaatgcaaggaaggtactgtagtga 400  
QY 421 ttggaagtggagattttggccaatccttgaccatcgactatagatggcgctacatg 480  
|||||  
Db 401 ttggaagtggagattttggccaatccttgaccatcgactatagatggcgctacatg 460  
QY 481 tggctataggaagttagaataccttaagtttgccttgatattttcccatggtgtagatg 540  
|||||  
Db 461 tggctataggaagttagaataccttaagtttgccttgatattttcccatggtgtagatg 520  
QY 541 tcaatcatatgaaatgagctcacaacaaacaaataatattgtgtcatcacagag 600  
|||||  
Db 521 tcaatcatatgaaatgagctcacaacaaacaaataatattgtgtcatcacagag 580  
QY 601 aacattatcctccctgtggagcttgagacatctgtgtgggtaaatccctgattgatg 660  
|||||  
Db 581 aacattatcctccctgtggagcttgagacatctgtgtgggtaaatccctgattgatg 640  
QY 661 tggagataacatgagagataaacacagatccagatccatgctgaatttggtctcat 720  
|||||  
Db 641 tggagataacatgagagataaacacagatccagatccatgctgaatttggtctcat 700  
QY 721 tattccagattcttattatgtaaaagattatattgtctcaagcttggcgactcagt 780  
|||||

Db	701	tattcccgatctcttgattggtcctaaagattaaagtgtgtctcagcttggtcgacttcagt	760
Qy	761	taagaacctaaagatgycacgcccgcaggtttatataatgycacaaacatattcaagcgcac	840
Db	761	tagagcctaagaatgycacgcccgcaggtttatataatgycacaaacatattcaagcgcac	820
Qy	841	aacaggtatttgaacctgccccgcaggttgtaatttcattccattgacttggatcccttat	900
Db	821	aacaggtatttgaaccttgcgcgcaggttgtaatttcattccattgacttggatcccttat	880
Qy	901	catacgccagagagatgtgaaatttacccttaagactcttctactctctggaagagccag	960
Db	881	catacgccagagagatgtgaaatttacccttaagactcttctactctctggaagagccag	940
Qy	961	tgtgtgtagcatagacttggccacatctttctccttattcctcttgtcacagatgtga	1020
Db	941	tgtgtgtagcatagacttggccacatctttctccttattcctcttgtcacagatgtga	1000
Qy	1021	ttcatcatatgcttagaagaacacacagatgactcttacaacaaatccctatagagattgta	1080
Db	1001	ttcatcatatgcttagaagaacacacagatgactcttacaacaaatccctatagagattgta	1060
Qy	1081	ataaaccttaccatagcttggccaattctgtctcctccctagtataccctgcagagttctc	1140
Db	1061	ataaaccttaccatagcttggccaattctgtctcctccctagtataccctgcagagttctc	1120
Qy	1141	tggcagctgcttaccacttatttaacgcccacagatagagagattccacacttgcttg	1200
Db	1121	tggcagctgcttaccacttatttaacgcccacagatagagagattccacacttgcttg	1180
Qy	1201	aaacctggttacagtgtagaanaacacagcttgagattactaaagttttcttcgtatgtcc	1260
Db	1181	aaacctggttacagtgtagaanaacacagcttgagattactaaagttttcttcgtatgtcc	1240
Qy	1261	atgttgcccttaagacctctgtcttaccgataggaagtcacagagatatgtttctccaca	1320
Db	1241	atgttgcccttaagacctctgtcttaccgataggaagtcacagagatatgtttctccaca	1300
Qy	1321	tggcttaccagcag	1334
Db	1301	tggcttaccagcag	1314
RESULT 10			
ABA45820			
ID	ABA45820	standard; DNA; 1953 BP.	
XX	AC	ABA45820;	
XX	AC		
DT	01-FEB-2002	(first entry)	
XX	XX		
DE	Human breast cell single exon nucleic acid probe #4515.		
XX			
KK-	Human: microarray; single exon probe; gene expression; breast;		
XX	disease; cancer; ss.		
XX			
OS	Homo sapiens.		
XX			
PM	W0200157271-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	30-JAN-2001; 2001WO-US00662.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
PA	(MOLE-) MOLECULAR DYNAMICS INC.		

XX	Penn SG,	Hanzel DK,	Chen W,	Rank DR;
PI	WPI; 2001-496933/54.			
DR				
XX				
XX				
PT	New spatially-addressable set of single exon nucleic acid probes,			
PT	useful for measuring gene expression in sample derived from human			
PR	breast, comprises number of single exon nucleic acid probes			
PS				
XX				
XX	Claim 1; SEQ ID NO 4515; 327pp + sequence listing; English.			
CC	The invention relates to a spatially-addressable set of single exon			
CC	nucleic acid probes for measuring gene expression in a sample derived			
CC	from human breast and BR 474 cells. The method involves contacting			
CC	the probes with a collection of detectably labeled nucleic acids			
CC	derived from mRNA of human breast, and then measuring the label			
CC	bound to each probe of the microarray. The probes are useful for			
CC	verifying the expression of regions of genomic DNA predicted to			
CC	encode proteins. They are useful for gene discovery, and for			
CC	determining predisposition and/or prognosing breast disease. Gene			
CC	expression analysis is useful for assessing the toxicity of chemical			
CC	agents on cells. The microarray of this invention presents a far greater			
CC	diversity of probes for measuring gene expression, with far less bias			
CC	than expressed sequence tag microarrays. The method is suitable for			
CC	rapid production of functional information from genomic sequence. The			
CC	presented sequence is a single exon nucleic acid probe of the invention.			
CC	Note: The sequence data for this patent did not form part of the			
CC	printed specification, but was obtained in electronic format directly			
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.			
XX				
XX				
SQ	Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;			
Query Match	.28.1%;	Score 530.6;	DB 22;	Length 1953;
Best Local Similarity	99.3%;	Prod. No. 6.5e-147;		
Matches 533;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
OY	280	aggaatattcttggatccttggaagtgtccgcatatcgaaatcaatcctatgatggaa	339	
Db	343	aggatatctcttggtgatcttggaaatgcgcgatatcaatgaaatccttatgatggaa	402	
OY	340	gccctaagaagccttagtgaaccttgttaacctaatgtgataaattgtalcaaanatgcaa	399	
Db	403	gccctaagaagccttagtgaacctttttaccctaagcataatgtalatcaaanatgcaa	462	
OY	400	ggaagtcactgttaagtgtgatgtgaagtgtgaatatttgcacaatcccttgacattgac	459	
Db	463	ggaagtcactgttaagtgtgatgtgaagtgtgaattttgcacaatcccttgacattgac	522	
OY	460	ttaattagatlgcgccatcatgtgycgcataagagaagtagaalccttaagtttgcttgat	519	
Db	523	ttattagatlgcgccatcatgtgycgcataagagaagtagaalccttaagtttgcttgat	582	
OY	520	ttttcttcatatgtttagatgtcactcatcatgaagaatgcttccaanaaacaaataaa	579	
Db	583	ttttcttcatatgtttagatgtcactcatcatgaagaatgcttccaanaaacaaataaa	642	
OY	580	tattgttgcatatacaagagaacattactccctgttggaaccttgagacattgcttg	639	
Db	643	tattgttgcatatacaagagaacattactccctgttggaaccttgagacattgcttg	702	
OY	640	tgggtaaaaatccctgaattatgtgtagcaataaactatagagataaacagatcaccaagatcca	699	
Db	703	tgggtaaaaatccctgaattatgtgtagcaataaactatagagataaacagatcaccaagatcca	762	
OY	700	atgctgaatatttggtctcaatbatccagatcctttgatgtgcaaanagatttaaagt	759	
Db	763	atgctgaatatttggtctcaatbatccagatcctttgatgtgcaaanagatttaaagt	822	
OY	760	tctagcttggaacttaagttagaactaagatgacagccggaagttatatat	816	
Db	823	tctagcttggaacttaagttagaactaagatgacagccggaagttatatat	879	



CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence: 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 28.1%; Score 530.6; DB 22; Length 1953;  
Best Local Similarity 99.3%; Pred. No. 6.5e-147;  
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 280 aggatattctgtgacatggaagtgccgatacgaagaatcctcatatgaggaa 339
    |||
Db 343 aggatattctgtgacatggaagtgccgatacgaagaatcctcatatgaggaa 402
QY 340 gccctaaagccttagtgaactgtttacctaattgacataatggtatcaaatgacaa 399
    |||
Db 403 gccctaaagccttagtgaactgtttacctaattgacataatggtatcaaatgacaa 462
QY 400 ggaaggtcactgtgagtgatggaagtgaagatttgcacaatccttgaccattcgac 459
    |||
Db 463 ggaaggtcactgtgagtgatggaagtgaagatttgcacaatccttgaccattcgac 522
QY 460 ttattagatcgagatcatatgtagtaagaatcctaagtttgcttgaat 519
    |||
Db 523 ttattagatcgagatcatatgtagtaagaatcctaagtttgcttgaat 582
QY 520 ttcttcctcatgtgtgatagtcactcatatgaagaatgtcctcacaaaaaataataa 579
    |||
Db 583 ttcttcctcatgtgtgatagtcactcatatgaagaatgtcctcacaaaaaataataa 642
QY 580 tatttgtgtatatacagaagaataattatcctccctgtgtgacctgaagatctgcttg 639
    |||
Db 643 tatttgtgtatatacagaagaataattatcctccctgtgtgacctgaagatctgcttg 702
QY 640 tgggttaaacccctgattgattgagcaataacatgagataaacaatcaccagaatcca 699
    |||
Db 703 tgggttaaacccctgattgattgagcaataacatgagataaacaatcaccagaatcca 762
QY 700 atgcgtgaataattggcttcattatccagattccttgattgtcaagaatttaatttg 759
    |||
Db 763 atgcgtgaataattggcttcattatccagattccttgattgtcaagaatttaatttg 822
QY 760 ttcgaagcttgagcaatcgaattgagcaatgagcaatgagcaatgagcaatgagcaat 816
    |||
Db 823 ttcgaagcttgagcaatcgaattgagcaatgagcaatgagcaatgagcaatgagcaat 879
```

RESULT 13  
AAK04514  
ID. AAK04514 standard; DNA: 1953 BP.  
XX  
AC AAK04514;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 4505.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00667.  
XX  
PF 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483446/52.  
XX  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX  
PS Example 4; SEQ ID NO: 4505; 650bp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

CC  
XX  
SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 28.1%; Score 530.6; DB 22; Length 1953;  
Best Local Similarity 99.3%; Pred. No. 6.5e-147;  
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 280 aggatattctgtgacatggaagtgccgatacgaagaatcctcatatgaggaa 339
    |||
Db 343 aggatattctgtgacatggaagtgccgatacgaagaatcctcatatgaggaa 402
QY 340 gccctaaagccttagtgaactgtttacctaattgacataatggtatcaaatgacaa 399
    |||
Db 403 gccctaaagccttagtgaactgtttacctaattgacataatggtatcaaatgacaa 462
QY 400 ggaaggtcactgtgagtgatggaagtgaagatttgcacaatccttgaccattcgac 459
    |||
Db 463 ggaaggtcactgtgagtgatggaagtgaagatttgcacaatccttgaccattcgac 522
QY 460 ttattagatcgagatcatatgtagtaagaatcctaagtttgcttgaat 519
    |||
Db 523 ttattagatcgagatcatatgtagtaagaatcctaagtttgcttgaat 582
QY 520 ttcttcctcatgtgtgatagtcactcatatgaagaatgtcctcacaaaaaataataa 579
    |||
Db 583 ttcttcctcatgtgtgatagtcactcatatgaagaatgtcctcacaaaaaataataa 642
QY 580 tatttgtgtatatacagaagaataattatcctccctgtgtgacctgaagatctgcttg 639
    |||
Db 643 tatttgtgtatatacagaagaataattatcctccctgtgtgacctgaagatctgcttg 702
QY 640 tgggttaaacccctgattgattgagcaataacatgagataaacaatcaccagaatcca 699
    |||
Db 703 tgggttaaacccctgattgattgagcaataacatgagataaacaatcaccagaatcca 762
QY 700 atgcgtgaataattggcttcattatccagattccttgattgtcaagaatttaatttg 759
    |||
Db 763 atgcgtgaataattggcttcattatccagattccttgattgtcaagaatttaatttg 822
QY 760 ttcgaagcttgagcaatcgaattgagcaatgagcaatgagcaatgagcaatgagcaat 816
    |||
Db 823 ttcgaagcttgagcaatcgaattgagcaatgagcaatgagcaatgagcaatgagcaat 879
```

RESULT 14  
AAK30016  
ID AAK30016 standard; DNA: 1953 BP.  
XX  
AC AAK30016;  
XX  
DT 06-NOV-2001 (first entry)

```
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 4573.
XX DE
XX DE Human: bone marrow expressed exon: gene expression analysis; probe:
XX DE microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 4573; 658bp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match      28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 280 aggatattcttgtagtcttgggaagtcgcgtatcatggaatcaatctatgtggaa 339
DB 343 aggatattcttgtagtcttgggaagtcgcgtatcatggaatcaatctatgtggaa 402
QY 340 gccctaaggccttagtgaactgttttaactaatgcatlaaattglatcaagaatgcaa 399
DB 403 gccctaaggccttagtgaactgttttaactaatgcatlaaattglatcaagaatgcaa 462
QY 400 ggaagtcacttagttagttagtgaagtggaatttgcacaaatcttgaccatcgac 459
DB 463 ggaagtcacttagttagttagtgaagtggaatttgcacaaatcttgaccatcgac 522
QY 460 ttattagatgagcgtatcatgtgtgcataggaagtaaaccttaagtgttcgtgaat 519
DB 523 ttattagatgagcgtatcatgtgtgcataggaagtaaaccttaagtgttcgtgaat 582
QY 520 ttttccatgtagtgaatgtagtcaatcatgaagaatgctctcaaaaaaacaatatata 579
DB 583 ttttccatgtagtgaatgtagtcaatcatgaagaatgctctcaaaaaaacaatatata 642
QY 580 tatttgtagtatacaagaacaatatatacctccttgtaggaacctgagaacatctgtcg 639
DB 643 tatttgtagtatacaagaacaatatatacctccttgtaggaacctgagaacatctgtcg 702
QY 640 tgggttaaatcctgattgtagtgaataacatgaggaataaacgagtaaccagaaatca 699
|||||
```

```
DB 703 tgggttaaatcctgattgtagtgaataacatgaggaataaacgagtaaccagaaatca 762
QY 700 atcgtaatatctggcttcatctccagatctcttgtagtgcgaagaattatgttg 759
DB 763 atcgtaatatctggcttcatctccagatctcttgtagtgcgaagaattatgttg 822
QY 760 tctcagcttggcactcagtttaggaacctaaagatgacgagcgaggttatatat 816
DB 823 tctcagcttggcactcagtttaggaacctaaagatgacgagcgaggttatatt 879

RESULT 15
AA114606
ID AA114606 standard; DNA; 1953 BP.
XX AC AA114606;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #4539 for gene expression analysis in human cervical cell sample.
XX DE Probe; human; microarray; gene expression; cervical epithelial cell;
XX DE cervical cancer; ss.
XX KW Homo sapiens.
XX OS
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 25; SEQ ID NO 4539; 487bp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/publ1shed_pct_sequences.
XX SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match      28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 280 aggatattcttgtagtcttgggaagtcgcgtatcatggaatcaatctatgtggaa 339
DB 343 aggatattcttgtagtcttgggaagtcgcgtatcatggaatcaatctatgtggaa 402
|||||
```

QY 340 gccctaagagcccttagtgaacatctgttaccataatgcatataatgatalcaagaatgcaa 399  
|||||  
Db 403 gccctaagagcccttagtgaacatcttcttaccataatgcatataatgatalcaagaatgcaa 462  
QY 400 ggaaggtcacctgtatgtgtgtgtgaagtgaagattttgccaatcccttgaccattgac 459  
|||||  
Db 463 ggaaggtcacctgtatgtgtgtgaagtgaagattttgccaatcccttgaccattgac 522  
QY 460 ttattagatgcgcatcatgttggtcatagagaatgagaatccctaagtttgtctctgaat 519  
|||||  
Db 523 ttattagatgcgcatcatgttggtcatagagaatgagaatccctaagtttgtctctgaat 582  
QY 520 ttttccctcatgtgtgtatgatagttcaatcaatgaagatgctcacaacaaataataa 579  
|||||  
Db 583 ttttccctcatgtgtgtatgatagttcaatcaatgaagatgctcacaacaaataataa 642  
QY 580 tattgtgtcatatacagagaacattatacttccctgtgtggacctgagacatctgttg 639  
|||||  
Db 643 tattgtgtcatatacagagaacattatacttccctgtgtggacctgagacatctgttg 702  
QY 640 tgggtaaaatccctgattgatatgtgagcaataacatgagataaacacagatccagaatcca 699  
|||||  
Db 703 tgggtaaaatccctgattgatatgtgagcaataacatgagataaacacagatccagaatcca 762  
QY 700 atgctgaataatttgcctcatlalttcccaagattcttggattgttcaaggatttaagtgtg 759  
|||||  
Db 763 atgctgaataatttgcctcatlalttcccaagattcttggattgttcaaggatttaagtgtg 822  
QY 760 tctcagcttgggcaacttcaagttaagacctaagatgccaagccggcaggttatatat 816  
|||||  
Db 823 tctcagcttgggcaacttcaagttaagacctaagatgccaagccggcaggttatatt 879

Search completed: May 9, 2002, 02:12:01  
Job time: 13401 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 22:21:20 ; Search time 54.46 Seconds  
(without alignments)  
8529.006 Million cell updates/sec

Title: US-09-802-520-2  
Perfect score: 1891  
Sequence: 1 ggggaagcagctgagtgctg.....gtcaattatcggtgtga 1891

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	27.6	521	US-09-323-873A-7	Sequence 7, Appli
2	476.4	25.2	1213	US-09-083-521-3	Sequence 3, Appli
3	425.8	22.5	444	US-09-030-607-214	Sequence 214, App
4	425.8	22.5	444	US-09-439-313-214	Sequence 214, App
5	363	19.2	366	US-09-030-607-215	Sequence 215, App
6	363	19.2	366	US-09-439-313-215	Sequence 215, App
7	310.2	16.4	322	US-09-323-873A-9	Sequence 9, Appli
8	306.8	16.2	328	US-09-030-607-212	Sequence 9, Appli
9	306.8	16.2	328	US-09-439-313-212	Sequence 212, App
10	290.4	15.4	1195	US-09-323-873A-1	Sequence 1, Appli
11	243.4	12.9	250	US-09-030-607-213	Sequence 213, App
12	243.4	12.9	250	US-09-439-313-213	Sequence 213, App
13	232.2	12.3	3627	US-09-323-873A-6	Sequence 6, Appli
14	125.4	6.6	592	US-09-439-313-342	Sequence 342, App
15	106	5.6	448	US-09-323-873A-11	Sequence 11, Appl
16	97.6	5.2	401	US-09-323-873A-12	Sequence 12, Appl
17	41.2	2.2	7218	US-08-232-463-14	Sequence 14, Appl
18	39	2.1	7218	US-08-232-463-14	Sequence 14, Appl
19	36.6	1.9	1219	PCR-US93-06251-11	Sequence 11, Appl
20	36.4	1.9	152331	US-09-128-155-16	Sequence 16, Appl
21	36	1.9	1474	PCR-US94-00545-19	Sequence 19, Appl
22	36	1.9	3478	US-08-396-479B-1	Sequence 1, Appli
23	36	1.9	3478	US-08-818-823-1	Sequence 1, Appli
24	35.6	1.9	684	US-08-998-416-672	Sequence 672, App
25	35.6	1.9	6453	US-08-306-691B-14	Sequence 14, Appl
26	35.6	1.9	6453	US-09-209-668-10	Sequence 10, Appl
27	35.6	1.9	6453	US-09-356-952-8	Sequence 8, Appli

28	34.8	1.8	4695	2	US-08-231-193A-57	Sequence 57, Appl
29	34.8	1.8	4695	2	US-08-486-273A-57	Sequence 57, Appl
30	34.8	1.8	4695	3	US-08-940-086A-57	Sequence 57, Appl
31	34.8	1.8	4695	4	US-08-940-035A-57	Sequence 57, Appl
32	34.6	1.8	5544	3	US-08-851-843A-68	Sequence 68, Appl
33	34.6	1.8	5544	4	US-08-974-549A-111	Sequence 111, Appl
34	34.6	1.8	5544	4	US-08-854-050-68	Sequence 68, Appl
35	34.6	1.8	5544	4	US-09-430-323-68	GENERAL INFORMA
36	34.6	1.8	5613	2	US-08-463-418-1	Sequence 1, Appli
37	34.2	1.8	1281	4	US-08-998-416-3	Sequence 3, Appli
38	34	1.8	669	1	US-08-463-115-17	Sequence 17, Appl
39	34	1.8	669	1	US-08-465-388-17	Sequence 17, Appl
40	34	1.8	2249	3	US-08-814-052-19	Sequence 19, Appl
41	34	1.8	2279	3	US-08-814-052-17	Sequence 17, Appl
42	34	1.8	2300	3	US-08-814-052-18	Sequence 18, Appl
43	34	1.8	3183	2	US-08-939-218A-1	Sequence 1, Appli
44	34	1.8	3187	5	PCR-US95-06815-1	Sequence 1, Appli
45	34	1.8	3192	1	US-08-706-037-26	Sequence 26, Appl

## ALIGNMENTS

```
RESULT 1
US-09-323-873A-7
; Sequence 7, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.160SU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 60/091,183
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-7

Query Match      27.6%; Score 521; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 1e-148;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1049 ttacttttcaaatccctcatagatgttgataaacccttccattagttccattac 1108
      1 ttacttttcaaatccctcatagatgttgataaacccttccattagttccattac 60
Db      1 ttacttttcaaatccctcatagatgttgataaacccttccattagttccattac 60

OY 1109 ttgtcttcctcatagatgttgatcttcgtgagctgcttaccattatcaagg 1168
      1 ttgtcttcctcatagatgttgatcttcgtgagctgcttaccattatcaagg 120
Db      1 ttgtcttcctcatagatgttgatcttcgtgagctgcttaccattatcaagg 120

OY 1169 caccagatagagatcttcacacttggttggaacctggttacaagttagaacaacagct 1228
      1 caccagatagagatcttcacacttggttggaacctggttacaagttagaacaacagct 180
Db      1 caccagatagagatcttcacacttggttggaacctggttacaagttagaacaacagct 180

OY 1229 tggattactaagttttttcttgctatgtgtccatgttgcttaagctctgcttaccgat 1288
      1 tggattactaagttttttcttgctatgtgtccatgttgcttaagctctgcttaccgat 240
Db      1 tggattactaagttttttcttgctatgtgtccatgttgcttaagctctgcttaccgat 240

OY 1289 gagaagtcagagatattgtttctcaacatgcttaccagcaggttcatgcaaatat 1348
```

Db 241 gagagagcagagagatacttgcctcaacatggtctatcagaaggtctacgaataat 300  
Qy 1349 tgaacctcttggaatgagaagaatttgagaattgaatgatatatctcttgcat 1408  
Db 301 tgaacctcttggaatgagaagaagtttgagaattgaatgatatatctcttgcat 360  
Qy 1409 aatgaccttggtctactctccctcctcgcagctactctatccctcagtgagaatgc 1468  
Db 361 aatgaccttggtctactctccctcctcgcagctactctatccctcagtgagaatgc 420  
Qy 1469 tttaactgagagaatctcatctttatctacatctgataatgcctcgtcgtcat 1528  
Db 421 tttaactgagagaatctcatctttatctacatctgataatgcctcgtcgtcat 480  
Qy 1529 aagctctccatgtttaattatgagtggaagaagcctt 1569  
Db 481 aagctctccatgtttaattatgagtggaagaagcctt 521

RESULT 2  
US-09-083-521-3  
: Sequence 3, Application US/09083521  
: Patent No. 6048970  
: GENERAL INFORMATION:  
: APPLICANT: Lal, Preeti  
: APPLICANT: Guegler, Karl J.  
: APPLICANT: Corley, Neil C.  
: TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
: STREET: 3174 PORTER DRIVE  
: CITY: PALO ALTO  
: STATE: CALIFORNIA  
: COUNTRY: USA  
: ZIP: 94304  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/083,521  
: FILING DATE: Herewith  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: CERRONE, MICHAEL C.  
: REGISTRATION NUMBER: 39,132  
: REFERENCE/DOCKET NUMBER: PF-0527 US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (650) 855-0555  
: TELEFAX: (650) 845-4166  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1213 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: LIBRARY: PROSTUT10  
: CLONE: 1691243  
US-09-083-521-3

Query Match 25.2%; Score 476.4; DB 3; Length 1213;  
Best Local Similarity 99.8%; Pred. No. 6.7e-135;  
Matches 477; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1172 caagtataagagattccaccttggttggaacctgttacagtgtagaagaagccttg 1231  
Db 1 CAAGTATAGAGATTCCACCTTGTTGGAAACTGGTTACAGTGTAGAAAACAGCTTGG 60

Qy 1232 attactaagttttctctcgtatgctccatggtcctacagcctcgtcctaccagatgag 1291  
Db 61 ATTACTAAGTTTTTCTTCGGCTATGGTCCATGTTGCCATACAGCCCTGCTTACCGATGAG 120  
Qy 1292 aaggtcaagagatattggttctcaacatggtctatcagaaggtctacgaatatga 1351  
Db 121 AAGGTCAAGAGATATTGTTCTCACATGGCTTATCAGCAGGTTCTATGCAAAATATTGA 180  
Qy 1352 aaactcttggaatgagaagaatttgagaattgaatgatatatctcttgacataat 1411  
Db 181 AAACCTTGGAATGAGAGAGAGATTGGAGAAATGAAATGATATCCTTGGCATATAT 240  
Qy 1412 gaccttggtctactctccctcctcgcagctactctatccctcagtgagaatgcctc 1471  
Db 241 GACCTTGGCTTACTTCCCTCCCTGCGCAGTCATCTATCCCTTCAGTGAACAATGCTTT 300  
Qy 1472 aaactgagagaatctgattatctacgtctacactggaatgatactgcctgcacataag 1531  
Db 301 AAACGTGAGAGAAATTCAGTTTATTCAGTCATCTGAGATATCTGCTGCTCATAG 360  
Qy 1532 tacttccatgtttaattatgagtggaagaagccttggagaagagctacagatt 1591  
Db 361 TACTTCCATGTTTATTTATGATGAGAAACGAGCTTTGAGAGAGATCTACAGATT 420  
Qy 1592 ttatacaccacaaacttgctctcgtcgtctgttgccctcaatgtgaattcgtggt 1649  
Db 421 TTATACACACCACAAACTTGTCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 478

RESULT 3  
US-09-030-607-214  
: Sequence 214, Application US/09030607  
: Patent No. 6262245  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, David C.  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS  
: NUMBER OF SEQUENCES: 224  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED AND BERRY LLP  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: WA  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/030,607  
: FILING DATE: 25-FEB-1998  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Makl, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.427C3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206) 682-6031  
: INFORMATION FOR SEQ ID NO: 214:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 444 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
US-09-030-607-214

Query Match 22.5%; Score 425.8; DB 4; Length 444;  
Best Local Similarity 98.6%; Pred. No. 8.7e-120;

Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
OY 688 acccagaatccaatgctgaatattgtgtcattatcccaattcttgattgtcaag 747
    |||||||
Db 1 acccagaatccaatgctgaatattgtgtcattatcccaattcttgattgtcaag 60
OY 748 gatttaagtgtctcagcttggcactcagttaggaacctaaagatgcagccgcag 807
    |||||||
Db 61 gatttaagtgtctcagcttggcactcagttaggaacctaaagatgcagccgcag 120
OY 808 ttatataatgcaacaatatccaatgcgcagacaggttatgtacattgcccgcagt 867
    |||||||
Db 121 ttatataatgcaacaatatccaatgcgcagacaggttatgtacattgcccgcagt 180
OY 868 tgaattcattccattgacttgagatcccttatccatcagccagagattgaaatttac 927
    |||||||
Db 181 tgaattcattccattgacttgagatcccttatccatcagccagagattgaaatttac 240
OY 928 ccctacgaactcttactctctgagagagcgagtggtgtagctataagcttggccaca- 986
    |||||||
Db 241 ccctacgaactcttactctctgagagagcgagtggtgtagctataagcttggccaca- 300
OY 987 ttttttcccttattcccttctgtcagagatgtgattccatcatabtgcagaaaccaag 1046
    |||||||
Db 301 ttttttcccttattcccttctgtcagagatgtgattccatcatabtgcagaaaccaag 360
OY 1047 agtgaactttacaataatccctatagagattgtgaataaacccttactagttgccatt 1106
    |||||||
Db 361 agtgaactttacaataatccctatagagattgtgaataaacccttactagttgccatt 420
OY 1107 acttgcctccctagatactc 1130
    |||||||
Db 421 acttgcctccctagatactc 444
```

RESULT 4  
US-09-439-313-214  
; Sequence 214, Application US/09439313  
; Patent No. 6329505

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan Louise  
APPLICANT: Jiang Yugu  
APPLICANT: Reed, Steven G.  
APPLICANT: Kalos, Michael  
APPLICANT: Fanger, Gary  
APPLICANT: Retter, Gary  
APPLICANT: Solk, John

APPLICANT: Day, Craig  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

FILE REFERENCE: 210121.427C9

CURRENT APPLICATION NUMBER: US/09/439,313

CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 575

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 214

LENGTH: 444

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(444)

OTHER INFORMATION: n = A,T,C or G

US-09-439-313-214

Query Match 22.5%; Score 425.8; DB 4; Length 444;  
Best Local Similarity 98.6%; Pred. No. 8.7e-120;  
Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
OY 688 acccagaatccaatgctgaatattgtgtcattatcccaattcttgattgtcaag 747
    |||||||
Db 1 acccagaatccaatgctgaatattgtgtcattatcccaattcttgattgtcaag 60
OY 748 gatttaagtgtctcagcttggcactcagttaggaacctaaagatgcagccgcag 807
    |||||||
Db 61 gatttaagtgtctcagcttggcactcagttaggaacctaaagatgcagccgcag 120
OY 808 ttatataatgcaacaatatccaatgcgcagacaggttatgtacattgcccgcagt 867
    |||||||
Db 121 ttatataatgcaacaatatccaatgcgcagacaggttatgtacattgcccgcagt 180
OY 868 tgaattcattccattgacttgagatcccttatccatcagccagagattgaaatttac 927
    |||||||
Db 181 tgaattcattccattgacttgagatcccttatccatcagccagagattgaaatttac 240
OY 928 ccctacgaactcttactctctgagagagcgagtggtgtagctataagcttggccaca- 986
    |||||||
Db 241 ccctacgaactcttactctctgagagagcgagtggtgtagctataagcttggccaca- 300
OY 987 ttttttcccttattcccttctgtcagagatgtgattccatcatabtgcagaaaccaag 1046
    |||||||
Db 301 ttttttcccttattcccttctgtcagagatgtgattccatcatabtgcagaaaccaag 360
OY 1047 agtgaactttacaataatccctatagagattgtgaataaacccttactagttgccatt 1106
    |||||||
Db 361 agtgaactttacaataatccctatagagattgtgaataaacccttactagttgccatt 420
OY 1107 acttgcctccctagatactc 1130
    |||||||
Db 421 acttgcctccctagatactc 444
```

RESULT 5  
US-09-030-607-215/c  
; Sequence 215, Application US/09030607  
; Patent No. 6262245

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS

NUMBER OF SEQUENCES: 224

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/030,607

FILING DATE: 25-FEB-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: MAKI, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.427C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 215:

SEQUENCE CHARACTERISTICS:

LENGTH: 366 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-09-030-607-215

```
Query Match      19.2%   Score 363;   DB 4;   Length 366;
Best Local Similarity 99.2%   Pred. No. 1e-100;
Matches 363;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;
```

OY	1167	ggcccaagatagagaattccaccttgatgttggaacctgtgttcaatggtatgaaaaaag	1226
	1167		
Db	366	ggcaccagatragagaatttccacctgtgttggaacctgtgttcaagatgaaaaaag	307
OY	1227	cttgatctactaagttttcttcctgcgtatgltccatgltgtccaaagcctctgtaccg	1286
	1227		
Db	306	cttgatattactaagtgtttttnttgcgtatggtccatggtccctacagcctcgtgtaccg	247
OY	1287	atggaagaagtcagagagataattgtttctcacaatggtcttaacgaagttcaatgcaat	1346
	1287		
Db	246	atggaagaagtcagagagatattgtttctcacaatggtcttaacgaagttcaatgcaat	187
OY	1347	atggaagaagtcagagagagaaagtttggaagaattgaaatgtatatactccctttgc	1406
	1347		
Db	186	atgtgaaagctcttggaatgaggaagaaagtttggaagaattgaaatgtatatactccctttgc	127
OY	1407	ataatgagccttgcttaacttcccccctcgagcaatctcatccctcagtgagcaat	1466
	1407		
Db	126	ataatgagccttgcttaacttcccccctcgagcaatctcatccctcagtgagcaat	67
OY	1467	gctttaaactctgagaagaattcagtttattatcagctctacaactgtgatatgtcgtctgc	1526
	1467		
Db	66	gctttaaactctgagaagaattcagtttattatcagctctacaactgtgatatgtcgtctgc	7
OY	1527	ataagt 1532	
Db	6	ataagt 1	

```

RESULT      6
US-09-439-313-215/c
? Sequence 215, Application US/09439313
? Patient NO. 6329505
? GENERAL INFORMATION:
? APPLICANT: Xu, Jiangchun
? APPLICANT: Dillon, Davin C.
? APPLICANT: Mitcham, Jennifer L.
? APPLICANT: Harlocker, Susan Louise
? APPLICANT: Jiang Yuyun
? APPLICANT: Reed, Steven G.
? APPLICANT: Kalos, Michael
? APPLICANT: Fanger, Gary
? APPLICANT: Retter, Mark
? APPLICANT: Solk, John
? APPLICANT: Day, Craig
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
? FILE REFERENCE: 210121.42709
? CURRENT APPLICATION NUMBER: US/09/439,313
? CURRENT FILING DATE: 1999-11-12
? NUMBER OF SEQ ID NOS: 575
? SOFTWARE: FASTSEQ for Windows Version 3.0
? SEQ ID NO 215
? LENGTH: 366
? TYPE: DNA
? ORGANISM: Homo sapien
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(366)
? OTHER INFORMATION: n = A,T,C or G
?S-09-439-313-215

```

Query Match	19.28;	Score 363;	DB 4;	Length 366;
Best Local Similarity	99.28;	Pred. No. 1e-100;		
Matches 363; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

OY	1167	ggcccaagataagagattccaccttggttggaacgtgttaacgtgtgaaacacag	1226
Db	366	GGCCACCAAGATAGAGATTCCACCTTGSTTGGAACCTGGTTACAGTGTGAAAAACAG	307
OY	1227	cttgatctactaagttttcttcctgcgtatgltccatgltgcctaaagcctctgttaacg	1286
Db	306	CTTGAGATTACTAAGTTTTTTNTTCGGTATAGTCCATGTTGCTACAGCCTGTGTACCG	247
OY	1287	atggaaggtcagagagataattggtttctcaacatgcttacgcaggttcaatgcaaat	1346
Db	246	ATGGAAGGTCAGAGAAATTTGTTTCTCAACATGGCTATCACAGGTTCAATGCAAAAT	187
OY	1347	attgaaacctcttggaatgsgaagaagtttgagaattgaaatgatatactcccttggc	1406
Db	186	ATTGAAAACTCTTGGAATGAGAGAAAGTTGGAGAAATGTAATATATATTCCTTGGC	127
OY	1407	ataatgagccttgctctactcttcctccctcggcagtcactctcatcccttcagtagcaat	1466
Db	126	ATATATGAGCCTTGCTTACTTCCCTCGTGGCAGTCACTTATCCCTTCAGTAGCAAT	67
OY	1467	gctttaaactcgtgagagaatctcagtttattatcagtcctacaacttgaatgtgcctcgtc	1526
Db	66	GCTTTAAACTGAGAGAAATCACTTTTATTACAGTATACACTGGATATGTCGCTCGTGC	7
OY	1527	ataagt 1532	
Db	6	ATAAGT 1	

RESULT 7  
 US-09-323-873A-9  
 Sequence 9, Application US/09323873A  
 Patent No. 6329503  
 GENERAL INFORMATION:  
 APPLICANT: Daniel E. Afar  
 APPLICANT: Renee S. Hubert  
 APPLICANT: Arthur Leong  
 APPLICANT: Arthur B. Raitano  
 APPLICANT: Douglas C. Saffran  
 APPLICANT: Steve Chappell Mitchell  
 TITLE OF INVENTION: NOVEL SEPTENTINE TRANSMEMBRANE ANTIGENS  
 TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF  
 FILE REFERENCE: 129.16USD2  
 CURRENT APPLICATION NUMBER: US/09/323, 873A  
 CURRENT FILING DATE: 1999-06-01  
 PRIOR APPLICATION NUMBER: 60/087,520  
 PRIOR FILING DATE: 1998-06-01  
 PRIOR APPLICATION NUMBER: 60/091,183  
 PRIOR FILING DATE: 1998-06-30  
 NUMBER OF SEQ ID NOS: 32  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9  
 LENGTH: 322  
 TYPE: DNA  
 ORGANISM: Homo Sapiens  
 US-09-323-873A-9

[illegible]

QY 1171 ccaaglatagagagattccacctgtgtgaaacctgtgtacagltgtagaaacagcttg 1230  
|||||  
DB 188 ccaaglatagagagattccacctgtgtgaaacctgtgtacagltgtagaaacagcttg 247  
QY 1231 gattactaagttcttctcgtcgtacagltggtccacagcctctgtctacagata 1290  
|||||  
DB 248 gattactaagttcttctcgtcgtacagltggtccacagcctctgtctacagata 307  
QY 1291 gaagtcacagagat 1305  
|||||  
DB 308 gaagtcacagagat 322

RESULT 8  
US-09-030-607-212  
; Sequence 212, Application US/09030607  
; Patent No. 6262245  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
; NUMBER OF SEQUENCES: 224  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/030,607  
; FILING DATE: 25-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.427C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 212:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 328 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-09-030-607-212

Query Match 16.2%; Score 306.8; DB 4; Length 328;  
Best Local Similarity 97.9%; Pred. No. 1.2e-83;  
Matches 319; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 689 cccaagaatccaatgctgaataattgctcattatccacagattccttattgttcaagg 748  
|||||  
DB 3 CCAAAAATCCATGCTGATATTTGGCTTCATTATCCCANNTCTTTATGTCTCAAAG 62  
QY 749 attaatgtgtcctcagcttggtgacattcagtaaggaacctaaagatgccagcgaggt 808  
|||||  
DB 63 ATTATATGTTCTCAGCTTGGGCACTTCAGTTAGGACCTTAGGATGCCAGCGGAGGT 122  
QY 809 ttatatgacgaacaatatccaagcgagacaagagttatgaactgtcccgcaagt 868  
|||||  
DB 123 TTATATATGACACACATATTTCAAGCGGACACAGGTTATGACTTGCCGCGAGTT 182  
QY 869 gaattcattcccatgtactgtgagatccttatcatcagcagaagatgtgaaattacc 928

DB 183 NATTTTCATCCCATGACTGGATGCTTATCATAGCCAGAGATTGMAATTTACC 242  
QY 929 cctagacacttactctctcgtgagagggccagtggtgtgtagcataagcttgccaca-t 987  
DB 243 CCTACNACTCTTACTCTCTGAGNAGGCCAGTGtGTAGTATTAAGCTTGCCACATT 302  
QY 988 ttcttcttattcctctgtcaga 1013  
|||||  
DB 303 TTTTTCCTTATTCCTTTGTGAGA 328

RESULT 9  
US-09-439-313-212  
; Sequence 212, Application US/09439313  
; Patent No. 6329505  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yugu  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C9  
; CURRENT APPLICATION NUMBER: US/09/439,313  
; NUMBER OF SEQ ID NOS: 575  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 212  
; LENGTH: 328  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(328)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-439-313-212

Query Match 16.2%; Score 306.8; DB 4; Length 328;  
Best Local Similarity 97.9%; Pred. No. 1.2e-83;  
Matches 319; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 689 cccaagaatccaatgctgaataattgctcattatccacagattccttattgttcaagg 748  
|||||  
DB 3 cccaagaatccaatgctgaataattgctcattatccacagattccttattgttcaagg 62  
QY 749 attaatgtgtcctcagcttggtgacattcagtaaggaacctaaagatgccagcgaggt 808  
|||||  
DB 63 attaatgtgtcctcagcttggtgacattcagtaaggaacctaaagatgccagcgaggt 122  
QY 809 ttatatgacgaacaatatccaagcgagacaagagttatgaactgtcccgcaagt 868  
|||||  
DB 123 ttatatgacgaacaatatccaagcgagacaagagttatgaactgtcccgcaagt 182  
QY 869 gaattcattcccatgtactgtgagatccttatcatcagcagaagatgtgaaattacc 928  
|||||  
DB 183 naattcattcccatgtactgtgagatccttatcatcagcagaagatgtgaaattacc 242  
QY 929 cctagacacttactctctcgtgagagggccagtggtgtgtagcataagcttgccaca-t 987  
DB 243 cctagacacttactctctcgtgagagggccagtggtgtgtagcataagcttgccacaatt 302  
QY 988 ttcttcttattcctctgtcaga 1013  
|||||  
DB 303 ttcttcttattcctctgtcaga 328

```

RESULT 10
US-09-323-873A-1
; Sequence 1, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappel Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129, 16USU2
; CURRENT APPLICATION NUMBER: US/09/323, 873A
; PRIOR FILING DATE: 1998-06-01
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1195
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-1

```

```

Query Match 15.4%; Score 290.4; DB 4; Length 1195;
Best Local Similarity 60.8%; Pred. No. 2.7e-78;
Matches 474; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

```

```

QY 935 acccttacccttgagagagggccagtggtgagctatagcttggccattttt 994
DB 263 acccttccacagtggaacttcaataaagctgtatatagcattcctgcatt 322
QY 995 ccttattccttgacagagatgtatcatcatatgtagaagaacagagtgact 1054
DB 323 tctttacacctcttgagaggaagtaatacctttagcaacttccatacaaat 382
QY 1055 ttacaataatcctatagagatgtgtaaaacctaactatagttgcaacttgc 1114
DB 383 ttataaataatccaactctgtgtacaaagaagcttgcacatgttccactcct 442
QY 1115 ctccctagatctacagagcttctctgagagctgtctacacattatagaag 1174
DB 443 ggcattggttaccctgcagagtgtagaagcaaatgtcccaacttcaataag 502
QY 1175 gtagagagatctccaccttggttggaacctgttaccagttagaanaacag 1234
DB 503 gtagaagaagttccacatggttgtagaagtgtagttagaagaagaagttg 562
QY 1235 actaagtttctctgcgtatggtccatgttgcctacagcctctgcttacc 1294
DB 563 tctcagtttcttcttctgtactgtactgtacgaattatagctgttaccac 622
QY 1295 gtagagagatatttttctcaacatggttctcagcaagttcagcaaat 1354
DB 623 atcctcagatatacagttgtaaacctggtgcatatcaacaggtccaaata 682
QY 1355 cctttagaagagagaaagtttggaagaatggaatgtatatccttggcaat 1414
DB 683 tgccttgatgagagatgatttggaagaatggaatgtatgtctctcgtgaa 742
QY 1415 ccttggcttacttccctcctgagcagctacttacccttccagtgagcaat 1474
DB 743 atgggaataactgctctgtgctgtgacatattccatctcgtgagtgacct 802
QY 1475 ctggagagaatcagtttatcagttcagttcagttgtagtgcctgtcatt 1534

```

```

DB 803 atggagagaattccactatatccagagcaagtaggaattgttcccttact 862
QY 1535 ttccacgtttaaattatgatgaaacagagctttagaagaagtagctacagat 1594
DB 863 aatacagcagatatttctgctcgtgaataagtgtagatataaacaatttga 922
QY 1595 taacacccaacttcttctgtctctgttctgcccataatgtaattctggt 1654
DB 923 tacaccccaactttagtagctgttcttccctcaatgtgtctcgtatatta 982
QY 1655 tatttattcctccatgataagccgaagcctaaacgaatataaagaagctg 1714
DB 983 catactattcctgcagatgcttgaggaagaactagatagatagatagatg 1042

```

```

RESULT 11
US-09-030-607-23/c
; Sequence 213, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaqun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030, 607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-213

```

```

Query Match 12.9%; Score 243.4; DB 4; Length 250;
Best Local Similarity 97.6%; Pred. No. 2e-64;
Matches 244; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1283 accgatgagaagtcagagagataattgttctcaacatggtctatcagcag 1342
DB 250 ACCGATGAGAAGGTMAGAGATATTGTTNTTACATGNTTATCAGCAGGTT 191
QY 1343 aaatatgaaaactcttggaagaagagatttggaagaatggaatgtatct 1402
DB 190 AAATATTGAAAACTCTTGGAATGAGGAAGATTGGAATGAAATGATATNT 131
QY 1403 tggcataaagagcttggtacttccctcctgagcagctacttacttccctc 1462
DB 130 TGGCATATAGCCTTGCTTACTTTCTCTCTGCGAGTCACCTTCTTCCCTT 71

```

OY	1463	caatgctttaaactggaggaatcagtttcttttaacgtacactggatgctgcct	1522
Db	70	CAATGCTTTAACTGGAGCAATTCGTTTATTCAGTCTAACTGNGATATGTCGCT	11
OY	1523	gctcaataagt	1532
Db	10	GCTCATTAAGT	1

RESULT 12

```

US-09-439-313-213/c
; Sequence 213, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42769
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ. ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(250)
; OTHER INFORMATION: n = A,T,C or G
; US-09-439-313-213

```

Query Match	12.9%	Score 243.4	DB 4	Length 250
Best Local Similarity	97.6%	Pred. No. 2e-64		
Matches 244	Conservative 0	Mismatches 6	Indels 0	Gaps 0
QY 1283	accgatgagaaggtcagaagatatcttgcttcacacatggtcttcacagcaggtcaltgc	1342		
Db 250	ACCCATGAGAGGTNACGAGATATTGTNTTAACTGGTATGNTATVCAGAGTTCATGC	191		
QY 1343	aaatctgaaaaactcttggaatggaaggaaggtttgagaattgaatgatatctctt	1402		
Db 190	AAATTTGAAAACCTTGGATGAGGAGAACTTTGGAGAAATTGAAATGTATATTCCTT	131		
QY 1403	tggcataatgagccttggcttactctccctccctggcagctcaacttcatcccttcagtag	1462		
Db 130	TGGCATATGTAGCCTTGGCTTACTTTCCTCCCTCGACACTCTTATCCCTTCAGTAG	71		
QY 1463	caatgctttaaactgggaagagaatcgaatttatcattcaagctcaacttgatatgtcgcct	1522		
Db 70	CAATGCTTTAAACGTGAGAGAAATTCAGTTTATTTCAGCTPACACACINGATATATGTGCTCT	11		
QY 1523	gctcataagt 1532			
Db 10	GCTCATAAGT 1			

## RESULT 13

US-09-323-873A-6  
; Sequence 6, Application US/09323873A  
; Patent No. 6329503  
; GENERAL INFORMATION:

```

: APPLICANT: Daniel E. Afar
: APPLICANT: Rene S. Hubert
: APPLICANT: Kahan Leong
: APPLICANT: Arthur B. Raltano
: APPLICANT: Douglas C. Saffran
: APPLICANT: Steve Chappell Mitchell
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
: FILE REFERENCE: 129,160S02
: CURRENT APPLICATION NUMBER: US/09/323,873A
: CURRENT FILING DATE: 1999-06-01
: PRIOR APPLICATION NUMBER: 60/087,520
: PRIOR FILING DATE: 1998-06-01
: PRIOR APPLICATION NUMBER: 60/091,183
: PRIOR FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 3627
: TYPE: DNA
: ORGANISM: Homo Sapiens
: US-09-323-873A-6

Query Match      12.3%   Score 232.2: DB 4:   Length 3627;
Best Local Similarity 63.2%:   Pred No. 2.9e+60;
Matches 357; Conservative 0; Mismatches 208; Indels 0; Gaps 0

```

## RESULT 14

US-09-439-313-342/c  
; Sequence 342, Application US/09439313  
; Patent No. 6329505

```

; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuyou
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-342

```

Query Match 6.6%; Score 125.4; DB 4; Length 592;

Best Local Similarity 61.7%; Pred. No. 2.8e-28; Matches 198; Conservative 1; Mismatches 122; Indels 0; Gaps 0;

```

QY 935 acccttactctctgagagagccagtggtgtagctataagcttggccacatttttt 994
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 321 ACCTTTCCACAGTGGCAGCTTGCCAAATTAATAAGCTGCTATTATAGCATCTGACTTT 262
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 995 ccttattccttctgcaagagatgtatcatcatatgtcagaaaccaaagagtgact 1054
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 261 TCTTTACACTCTCTGAGGAGATTAATTCACCTTTAGCACTTCCATCAATATTTT 202
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1055 ttacaactctctatagagatgtgtaataaacttaccatattgtgcatcttctgc 1114
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 201 TTATTAATATTCATCTGCTGTCATCAACAAAGCTTGCCAAATGTTCCATCACTGCT 142
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1115 ctccctagatattctcgaaggtcttctgcaagctgttatacaacttatttaagcacc 1174
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 141 GGCATTGGTTTACTGCTGACAGTGTGATGACGCAATGTGCAACTTCATATGGAACAA 82
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1175 gttaggagatttccacccttggttggaacacttggtacagtgtaaaaacagcttgatt 1234
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81 GATTAAGAAAGTTTCACATTTGGTTGGATGATGTTAACAAGAACGARAATTGGGCT 22
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1235 actaagtttttcttcgcat 1255
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21 TCTCAGTTCTTTTGTGCTGT 1
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15  
US-09-323-873A-11

; Sequence 11, Application US/09323873A

; Patent No. 6329503

; GENERAL INFORMATION:

; APPLICANT: Daniel E. Afar

; APPLICANT: Rene S. Hubert

; APPLICANT: Kahan Leong

; APPLICANT: Arthur B. Raitano

; APPLICANT: Douglas C. Saffran

; APPLICANT: Steve Chappell Mitchell

; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS

; FILE REFERENCE: 129,160SU2

; CURRENT APPLICATION NUMBER: US/09/323,873A

; CURRENT FILING DATE: 1999-06-01

; PRIOR APPLICATION NUMBER: 60/087,520

; PRIOR FILING DATE: 1998-06-01

```

; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-09-323-873A-11

```

Query Match 5.6%; Score 106; DB 4; Length 448;

Best Local Similarity 56.6%; Pred. No. 1.9e-22; Matches 196; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

QY 1373 agttggagatggaatgatatctcttggcataatgagcttggcttccct 1432
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 agcttgctcagtgatcatatgtgcttgggaataactgggttttcttctgttgc 89
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1433 cctggcagtcactctatccctcagtgagcaatgcttaactggagagaattcagtt 1492
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 90 ctgggaatacactcttggccatctgttgaagcaatgcaactggagagattccgatt 149
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1493 taltcagttcaacttggatattgtctctgctccataagtaacttccatttattta 1552
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 tglccagttccaacttgggtatttlttgaccctgactctgtgtacagccacaccc 209
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1553 tggatggaagagagcttltgaggaagagtaactacagatttatacaccacaaattgt 1612
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 cggltgggaagagattctcagcccttcaatctcagatgtaacttctcctcagc 269
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1613 tcttgctctgttlttgccctcaattgtaattctgggtgaagattatttattcttcc 1672
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 270 gttagggttattcattcttgcactgtgctgtgattcaagttgttccttaatactg 329
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1673 tataagccgaagactaaacgaattaaaagagcttgggaagaagc 1718
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 330 tglagacaacaccccttacaagagatccgcaggtctgggaagaagc 375
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: May 9, 2002, 02:06:15  
Job time: 13495 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 20:23:30 ; Search time 2036.92 seconds  
(without alignments)  
12530.066 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagctgagtcgctgcg.....gtcaattatcgtgtgtta 1891

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Typal number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524.6	27.7	595	10	B1359597 384188 MA
2	451	23.8	561	10	B1133492 UI-M-BH3-
3	433.2	22.9	483	10	BF883142 QV3-ET021
4	411	21.7	712	12	A2428664 IM0212F14
5	377.8	20.0	2473	11	AK015015 Mus muscu
6	366	19.4	558	10	BM431438 B1105662
7	343.2	18.1	677	10	BM431438 ID0216F03
8	326.6	17.3	948	10	BF784438 B1105662
9	321.8	17.0	1287	11	AK003108 Mus muscu
10	310.2	16.4	322	9	AA508880 ng86f03.s
11	275.4	14.6	573	10	BM491178 pgp2n.pk0
12	274.8	14.5	541	9	A1747886 ui03e11.y
13	272.8	14.4	1211	11	AK010437 Mus muscu
14	262.2	13.9	680	9	BB603988 BB603988
15	260.8	13.8	343	9	AA957889 UI-R-E1-f
16	250.6	13.3	600	10	B1360480 387409 MA
17	237.4	12.6	881	10	BS655247 602582917

18	235	12.4	506	9	A1930004
19	232.4	12.3	1035	9	AL555520
20	224	11.8	379	9	A1930312
21	211	11.2	625	10	B6712772
22	209.8	11.1	950	10	C23400
23	207.4	11.0	336	10	BF560333
24	199.6	10.5	691	10	BS919620
25	199	10.5	675	10	BE380978
26	194.4	10.3	444	9	BB859054
27	190.4	10.1	255	9	AA957846
28	186.4	9.9	710	10	BB875216
29	185.2	9.8	1003	10	B1551229
30	182	9.6	959	9	AL515863
31	180.6	9.6	242	9	AA997082
32	177.8	9.4	1073	9	AL577999
33	176.6	9.3	920	10	BB881257
34	176.2	9.2	654	10	B1066339
35	174.8	9.2	742	9	AU121735
36	174.2	9.2	734	10	BC923646
37	169	8.9	694	10	BC918803
38	169	8.9	796	10	BG182296
39	167	8.8	521	10	BF022394
40	166	8.8	600	9	A1686139
41	165.6	8.8	513	9	AA032221
42	164.4	8.7	961	10	BG762026
43	163.6	8.7	464	9	AW159609
44	158.6	8.4	756	10	B1753475
45	157.4	8.3	896	10	B1149488

## ALIGNMENTS

RESULT 1  
BI359597 LOCUS 595 bp mRNA linear EST 01-AUG-2001  
384188 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.  
BI359597  
BI359597.1 GI:15055625

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

source  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACGCTGACCAT  
BACKWARD: GTTTCCAGTCACGACG  
Plate: 132 row: H column: 3  
Seq primer: ATTAGTCACACTATAG.  
Location/Qualifiers  
1..595  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 2P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;



QY 673 tgaagataaaccagtaaccagaatccaatgctgaatatttgcttcatattatccagatt 732  
|||||  
Db 381 TGAGACTAACCAGTACCAGATCCAGATGACAGAGTACCTGGCTCATATTATCCGACT 322  
QY 733 ctctgatttcaagaatttaattgctcagcttgaggcaactcagtagaagcctaag 792  
|||||  
Db 321 CCTGTATGTCAGAAAGATTATATGTCATCTGAGCTTGAGGACACTTACAGTAGGTCACAG 262  
QY 793 atgacgacgaggtttatatgacgaacaatatctaagggagcaacaggttatg 852  
|||||  
Db 261 ATGCCAGCCGCCAGGTTTATATATGCAACAAATATCCAGCTCCAGCAGAGTTATG 202  
QY 853 aacttccgacagttgaatttccatccatgacttgatgacttcatatcagccagag 912  
|||||  
Db 201 AGCTGCCGCCAGCGATTTTATTCCTGTTGACTTGAGATCTTGTGTCAGCCAG 142  
QY 913 agattgaataattaccctcagactcttactctctgagaggccagtggtgtagcta 972  
|||||  
Db 141 AGATTGAAACCTTACTCTGCGACTGTTACTCTGAGAGGGCCAGTGTAGTACCA 82  
QY 973 taagcttgacacatttttttcttcttcttcttcttcttcttcttcttcttctt 1032  
|||||  
Db 81 TAAAGCTTGCCACATTTTCTTCTTATTCCTTGTGCAAGATGTGATACATCATATG 22  
QY 1033 ctgaacaacacaga 1047  
|||||  
Db 21 CCAGAAAAA 7

RESULT 3  
LOCUS BF883142 483 bp mRNA linear EST 17-JAN-2001  
DEFINITION OV3-ET0211-071200-529-b10 ET0211 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BF883142  
VERSION BF883142.1 GI:12273268.  
KEYWORDS EST.  
SOURCE human.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 483)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

TITLE Sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2706922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&t2=QV3-ET0211-  
071200-529-b10&t3=2000-12-07&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 26  
High quality sequence stop: 483.

FEATURES  
source 1. 483  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="ET0211"  
/dev\_stage="Adult"

/note="Organ: lung-tumor; Vector: puc18; Site\_1: Smal;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 144 a 100 c 99 g 140 t  
ORIGIN

Query Match 22.9%; Score 433.2; DB 10; Length 483;  
Best Local Similarity 99.1%; Pred. No. 1.8e-99;  
Matches 446; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 462 attagatgggctcatatgctgctcaatgaagaatccatgcttcttgaatt 521  
|||||  
Db 483 ATTAGATGGGCTCATATGCTCATGAGAAATCTAAAGTTCTCTGAAATT 424  
QY 522 ttccctcatggtgtagatgtcactcatcgaagaatgctctcaaaaacaataata 581  
|||||  
Db 423 TTTCTCATGTGTGATGATGTCTCATGAGATGCTCTCAGAAAAAATATATTA 364  
QY 582 ttgtgtctatacacagagaacatlaaccctcttggagacctgagacatgctgtg 641  
|||||  
Db 363 TTGTGTGATACACAGAGAACATTAACCTCTGAGAGACTGAGACATGCTGTG 304  
QY 642 ggtataaatccatgattgattgagcaataacagagagataaacagatcccaatccat 701  
|||||  
Db 303 GGTAAAAATCCTGATGATGAGCAATTAACATGAGATTAACACAGTACCAGATCCAA 244  
QY 702 gctgaatattggtctcatattccgaattcttgattgtcaagaatttaattgtgtc 761  
|||||  
Db 243 GCTGAATATTGCTCTCATATTATCCGATTCCTTTATTGTCAAGGATTATATGTTG 184  
QY 762 tcaagctggacactcagttgaagaaatgacagcagcagagttatataatgcagc 821  
|||||  
Db 183 TCAGCTTGGGCACTTACGTAGAGACTTAAGATGCCAGCGGAGGTTATATATGAC 124  
QY 822 aacaatatcaagcgcagacaacaggtatgtaactgcccagctgtaattcattcc 881  
|||||  
Db 123 AACCAATATTCAGCGCGAGAACAGGTTATGAGCTTGCCGCCAGTGAATTCATCC 64  
QY 882 at-tgacttgagatctctatcctcagcag 910  
|||||  
Db 63 ATGTGACTTGGAGATCTTATATCATCAGAG 34

RESULT 4  
LOCUS AZ428664 712 bp DNA linear GSS 03-OCT-2000  
DEFINITION IM0212F1AF Mouse 10kb plasmid U00C1M library Mus musculus genomic  
clone U00C1M0212F14 F, DNA sequence.  
ACCESSION AZ428664  
VERSION AZ428664.1 GI:10552677  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 712)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.,  
and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA



Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)

Db 121 TTTAAGAGAGTACTACAGGTTTATACACCAACCTTGTCTTGGCTTGTG 180  
QY 1629 ccctcaatctgaatcttggaagaatattatcttccatgataaagccgaagct 1688  
Db 181 CCCCAATTTGAATTCGGGCAAGATGTTTACTCCCTCCATGATNAGCAAAAGCTA 240  
QY 1689 aaacgaatataaaagctgggaaagagccaaattctggaagaaggtattggaagca 1748  
Db 241 AAGGAATTAAAAAGGCTGGGAAAGGCCAATTTCTAGAAAGAGATTTGAGAGACA 300  
QY 1749 attccatgctcccccggagagaggtcacagtaatgataataatggtttcacagct 1808  
Db 301 GTTCCATCTCTACCGGAGAGGCTTACAGTAATGATATAAATGAGTCTCCCTGCT 360  
QY 1809 gccataaagttactacatgcatattttatagactc---taagctagttacaag 1865  
Db 361 GCTACATGAAGTCTCCGATGCGCATATTTTATGACTTCTTACTTCTTCACTTGCAAG 420  
QY 1866 tatgctgcaaatatcgtggttga 1891  
Db 421 TGCTCTGTCAAACTGCTGGGCTGA 446

RESULT 7  
LOCUS B1105662 677 bp mRNA linear EST 26-JUN-2001  
DEFINITION 602892429F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5037414 5',  
mRNA sequence.  
ACCESSION B1105662  
VERSION B1105662.1 GI:14556555  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 677)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM1103 row: j column: 07  
High quality sequence stop: 674.  
Location/Qualifiers  
1. 677

FEATURES  
source

/organism="Mus musculus"  
/strain="C2BEH II (feral)"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5037414"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
Stem cell origin."  
/lab\_host="DH10B"  
/note="Organ: Lung; Vector: PCMV-SPORT6; Site: 1: Salt;  
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 177 a 172 c 160 g 168 t  
ORIGIN

Query Match 18.1%; Score 343.2; DB 10; Length 677;  
Best local Similarity 85.6%; Pred. No. 1.8e-76;  
Matches 440; Conservative 0; Mismatches 68; Indels 6; Gaps 5;

QY 280 aggaattcttgatgacttggaagtgatcgatcagatgaatcaatctctatgtggaa 339  
Db 166 AGGTCATTCCGGGCTGCTTAGAGAGCTCTTACAGGAATTCATGTATGGAA 225  
QY 340 gccctaagagccttagtgaacttgcttaacctaatgataatgata-tcaagaagca 398  
Db 226 GCCCTTAAGAGCTT--GGAGAGCTTTTGGCTTAATGGCATATGTAATCAAGAGCA 283  
QY 399 aggaaggtcacatgtaggtgtagtgaagtagtatttgcacaaatcttgaaccatcga 458  
Db 284 AGCAGAGTCACCGTGGGGGTATAGGAAGTGGGATTTTGGCAAGTCTGTACCATTCGG 343  
QY 459 ctattagatgagctatcatatggtcataggaagtgaatcctaagtgtcttcgaa 518  
Db 344 CTATATGATGTGCTACACAGTGTGATAGGAAGCAAGAAATCCCAAGTTGCAATCAGA 403  
QY 519 tttttctcactggtgtagatgtagtccatcatgaatgagctctcacaaaacaatata 578  
Db 404 TTTTTCCTCACGTGTAGACGTACCCACCATGAAGTCTTTAACAAAATAATA 463  
QY 579 atattgtctatacacagagaacattataccctccgttggaacctgagacatgctt 638  
Db 464 ATATTGCGGCTATCCATAGAGAACATTTACACCTCTTGAGGACCTTAGACATCTGCT 523  
QY 639 gtggtataaatccatgattgtagtgaacaataacatgaggaataacagatccagaa-tc 697  
Db 524 GTGGGCAAAATCCATGATGATGAGCAACATGAGAGTAACAGTACCCAGAAATTC 583  
QY 698 caatgcgaatatttgcttcattatccagatcttctgattgataaagatttaa-tg 756  
Db 584 CAATGCAGAGTACTGCTTCATTTATTCGCCGACTCTTGATGTGTAAGAGATTAACTG 643  
QY 757 ttgtctagcttg-ggcactcagttagaccta 789  
Db 644 TGATCTCAGCTTGCGGCGCTTACACTAGTGTCCCA 677

RESULT 8  
LOCUS BF784438 948 bp mRNA linear EST 12-JAN-2001  
DEFINITION 602110826F1 NCI\_CGAP\_K1d14 Mus musculus cDNA clone IMAGE:4238935  
5', mRNA sequence.  
ACCESSION BF784438  
VERSION BF784438.1 GI:12089474  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 948)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9851 row: d column: 08  
High quality sequence stop: 660.  
Location/Qualifiers  
1. 948

FEATURES  
source

/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4238935"  
/clone\_lib="NCI\_CGAP\_K1d14"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Kidney; Vector: PCMV-SPORT6; Site: 1: NotI;

Site-2: Salt; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library. |"  
 BASE COUNT 228 a 244 c 231 g 245 t  
 ORIGIN

Query Match 17.3%; Score 326.6; DB 10; Length 948;  
 Best Local Similarity 82.6%; Pred. No. 3.4e-72;

Matches 422; Conservative 0; Mismatches 84; Indels 5; Gaps 4;

QY 1330 agcagggttcgaataattgaataactcttggaatgagaagaatttggaattgaa 1389  
 |||||  
 DB 11 AGCAGGTCATGCCAATATTGAGACGCCGGAACGAGGAGGAGTCTGGAGATTGAGA 70  
 |||||  
 QY 1390 tgcatactcccttggaataatgagccttggtctactctccctccggaagtcactct 1449  
 |||||  
 DB 71 TGTACATTTCC-TTGGCATCATAGGCTGGGCTT-GTGTCCCTGCGGCGTCACTTCCA 128  
 |||||  
 QY 1450 tcccttcagtgagcaatgctttaactggaagaatcttcttctacgtctacacttg 1509  
 |||||  
 DB 129 TCCCATCATGAGCAACGCTTTGAAGAGAGAGAGTTCAGTTCATTCAGCTTACGCTTG 188  
 |||||  
 QY 1510 gataatgcctctgctcacaatgacttccatgtttaaattatgagtggaacagactt 1569  
 |||||  
 DB 189 GCTACGCTCCGCCCTGCTCATACAGACCTTCACGTTTAATTACGATGGAACGCTGC-C 247  
 |||||  
 QY 1570 ttgagaagaagatgactcagattttataaccaccaacttgccttgcctgtctgttgc 1629  
 |||||  
 DB 248 TTCCACAAAGAGTACACCCCTTTTACACACCCAACTTCGTTGCTGCTGTTGCC 307  
 |||||  
 QY 1630 cctcaattgtaattctggaatgaattatttcttctcctcactgataagccgaagctaa 1689  
 |||||  
 DB 308 CTCAC--TGTAAVTTCGGGAAGATGATATTACTCTCCATCATTAAGCGGAACCTAA 365  
 |||||  
 QY 1690 aacgaattaaaaaaggctgggaagaagcaattcttggaagaaggtattggaagaa 1749  
 |||||  
 DB 366 AAGCAATTTAAAGGGCTGGGAAGAGGCCAGTTCTAGACGAGCATGGGAGGAGCGG 425  
 |||||  
 QY 1750 ttccatcttcccccggagagaggtacagatgattgagatgaataagtggtttacagctg 1809  
 |||||  
 DB 426 TTCTCATATCTGTCCTCCAGAGAGGGTCCACAGTGTGATGACAAAGAGGTCTCGCAAG 485  
 |||||  
 QY 1810 ccataataagtttactactatgcatatttt 1840  
 |||||  
 DB 486 CCACACACACTTCCATTCATTCATTCATTACTT 516  
 |||||  
 RESULT 9  
 AK003108 1287 bp mRNA linear HTC 19-JAN-2002  
 LOCUS  
 DEFINITION Mus musculus adult male heart cDNA, RIKEN full-length enriched  
 library, clone:1010001D01:homolog to CDNA FJ110829 FIS, CLONE  
 NT28P400138, full insert sequence.  
 ACCESSION AK003108  
 VERSION AK003108.1 GI:12833560  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (strain:C57BL/6J) adult male heart cDNA to mRNA,  
 clone:1010001D01.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (sites)  
 AUTHORS Carninci,P. and Hayashizaki,Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2 (sites)  
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
 Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

## TITLE

JOURNAL  
 MEDLINE  
 PUBMED

## REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

## COMMENT

Prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159  
 3 (sites)  
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
 Kono,H., Akiyama,J., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M.,  
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
 Fujiwaka,S., Inoue,K., Togawa,Y., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,  
 Yoneda,Y., Ishikawa,T., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN Integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multichannel sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861  
 4 (sites)  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 (bases 1 to 1287)  
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
 Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,  
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,  
 Hara,A., Hayatsu,N., Hill,D., Himoto,K., Hirooka,T., Hori,F.,  
 Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,  
 Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S.,  
 Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,  
 Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,  
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,  
 Schriml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,  
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
 Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamashita,I.,  
 Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and  
 Hayashizaki,Y.  
 Direct Submission  
 Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5'-GAGGAGAGAGAGGCGCCCACTCGAGTTTCTTTTCTTTTATV 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. cDNA went  
 through one round of normalization to Rot = 5.0. Second strand cDNA  
 was prepared with the primer adapter of sequence [5'-  
 GAGGAGAGAGAGTCCAGAGCTCATTTATTTATTAACCCCCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

## FEATURES

## source

Location/Qualifiers  
 1..1287  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGI:1900339"  
 /db\_xref="taxon:10090"  
 /clone="1010001D01"  
 /sex="male"  
 /tissue\_type="heart"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 misc\_feature  
 1..1287  
 /note="data source:SPRR, source key:G9NVB5, evidence:ISS







Qy 1111 tgcctccctagatatacctgcaggtcttcgtgcagctgcttatcaacttattacgca 1170  
|||||  
Db 128 tgcctccctagatatacctgcaggtcttcgtgcagctgcttatcaacttattacgca 187  
|||||  
Qy 1171 ccaagatagaagatccacacttggtgaaacctggttaccagtgtagaanaacagcttg 1230  
|||||  
Db 188 CCAAGTATAGAGATTTCACCTTGTTGGAACCTGGTTACAGTGTAGAAAACAGCTTG 247  
|||||  
Qy 1231 gattactaagttttcttcgtctatgtgtccatgtgtgctaccagctctgcttaccgata 1290  
|||||  
Db 248 GATTCTAGTGTCTTCCTGCTATGTCATGTCCTACAGCTCTGCTTACCGATGA 307  
|||||  
Qy 1291 gaagtcagagagat 1305  
|||||  
Db 308 GAGGTCAGAGAGAT 322  
|||||  
RESULT 11  
BM491178 573 bp mRNA linear EST 07-FEB-2002  
LACUS  
DEFINITION ppg2n.pk005.18 Normalized Chicken Pituitary/Hypothalamus/Pineal  
Library (ppg2n) Gallus gallus cDNA clone ppg2n.pk005.18 5' similar  
to gb|AAK50539.1 (AY029586) dudulin 2 [Mus musculus], mRNA  
sequence.  
ACCESSION BM491178  
VERSION BM491178  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 573)  
AUTHORS Porter,T.E. and Cogburn,L.A.  
TITLES ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA  
Library, USDA/IRAFs Animal Genome Project  
UNPUBLISHED (2002)  
JOURNAL Contact: Larry A. Cogburn  
COMMENT University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cogburnudel.edu, www.chickest.udel.edu.  
FEATURES  
Source  
1. 573  
/organism="Gallus gallus"  
/strain="Commercial broiler chickens"  
/db\_xref="taxon:9031"  
/clone="ppg2n.pk005.18"  
/library="Normalized Chicken  
Pituitary/Hypothalamus/Pineal Library (ppg2n)"  
/sex="Male and female"  
/tissue\_type="Pituitary Gland/Hypothalamus/Pineal Gland"  
/dev\_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9  
weeks)"  
/lab\_host="E. coli EMDH10B"  
/note="Vector: pCMVSPORT6; Library made from equivalent  
pools of total RNA isolated from each tissue at different  
ages. Single pass sequencing from 5'-end"  
BASE COUNT 150 a 153 c 144 g 126 t  
ORIGIN  
Query Match 14.6%; Score 275.4; DB 10; Length 573;  
Best Local Similarity 78.4%; Pred. No. 3.3e-59;  
Matches 330; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
Qy 298 ttggaagtcgcgtatcgtgaataatcctatgatgggaagccctaaagcctatgt 357  
|||||  
Db 152 TTCGGTCTCGAAGATGAGATCAATCTCATATATGGAAGCCCTAAGACCTCAATG 211  
|||||  
Qy 358 aaactgttacctaatgcatataatgatatcaagaagtcacactgtagtg 417  
|||||

Db 212 AAACCTTTCTACCAATATGAGCCATGAGCATGAGATGCCAGTAGATCACAGTAGGCA 271  
|||||  
Qy 418 tgattggaagtgagatatttcgcaaatccttgacatcctgacttattagatgcgtatc 477  
|||||  
Db 272 TCATTGGAAGTGGAGACCTTTGCAAAATCATTTGACCATTTGGGTGATCAATGTTGATCC 331  
|||||  
Qy 478 atgtgctcagtagaagtagaataatccaaagtgtgcttcgtaattttctcctcagtgtag 537  
|||||  
Db 332 ATGTTGTTGTAGAGACCAAAACCCCAAACTTGCTGACAGATTTCTTCCCATGtgtag 391  
|||||  
Qy 538 atgtcactcatatgaagaatgctctcacaacaaacaaataatattgtgtatcacaca 597  
|||||  
Db 392 ACGTCACATACCATGAGATGCGGTAGACAAACAAACATCTTTTGTAGCATATTCATA 451  
|||||  
Qy 598 gagaacattatacctccctgtgtgagcctgagacatcctgcttgggttaaatccgattg 657  
|||||  
Db 452 GAGAACACTACACGCTTTGTGGGACCTTCACGATTTACTCGGTAAAAATTCGTTG 511  
|||||  
Qy 658 atgtgagcaataacatgagataaaccagctaccacgaatccaatgctgataatttgctt 717  
|||||  
Db 512 ATGTACGACACATACAGAGTAGACCATATTCAGACTTCATGCAATGAGATTTGGCAT 571  
|||||  
Qy 718 c 718  
Db 572 c 572  
RESULT 12  
A1747886 541 bp mRNA linear EST 22-JUN-1999  
LOCUS A1747886  
DEFINITION ul03ell.y1 Sugano mouse kidney mRKA Mus musculus cDNA clone  
IMAGE:2065004 5', mRNA sequence.  
ACCESSION A1747886  
VERSION A1747886  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 541)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
JOURNAL The WashU-NCI Mouse EST Project 1999  
COMMENT UNPUBLISHED (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGT:994192  
Seq primer: custom primer used  
High quality sequence stop: 498.  
FEATURES  
Source  
1. 541  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2065004"  
/library="Sugano mouse kidney mRKA"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pME18S-FL3; Site\_1: DraIII  
(CACTGCTG); Site\_2: DraIII (CACTGCTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
(ATGTGCGCTTTTATTTTATTTT); double-stranded cDNA was



```

evidence:ISS
putative
six transmembrane epithelial antigen of the prostate"
/codon_start=1
/protein_id="BAB26938.1"
/db_xref="gi:12845881"
/translation="MEISDVTNPEQIMKMKPKNLIEDSDYSTDSGETSMKLRGLS
HLOAHVHADPCSELOHTEFFPNMRLPYKVAIIISLFEYLLEIPLYTSR
KQFYKPIIIVINVLPMVATLIALVLPGLIAVVOIANGSTYKRPPLDPMILA
KKQFGLISFFPNAVLAHAYSSISPMRSTRRLIMKAYKOYONKEDANVEDVPMEL
IVSLGIVGLATLALAVTISPSVSDSLTWRFHYIQSKLGLIVSLGLVHALVPMWK
WVDVSGFVWYMPPEFMAVFLPLVLICKIALCPILKKKILIKRCGWEDVSKINRTE
MASRL"
polyA_signal
1186..1191
/note="putative"
polyA_site
1211
/note="putative"
BASE COUNT      306 a      275 c      293 g      337 t
ORIGIN

```

```

Query Match      14.4%; Score 272.8; DB 11; Length 1211;
Best Local Similarity 59.4%; Pred. No. 1.9e-58;
Matches 463; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

```

```

OY 935 actcttactctctgagagagcagtggtgtagctaaagcttggccacattttttt 994
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 295 ATTCTTCCAAAGCTGGCGCTTCCAGTGAAGTCCCTGCATCATATCTATCCCTACCTT 354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 995 ccttattccttctgcagagatgtagtcatcatagctagaacaacacagagtgact 1054
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 335 CCTGTACACACTTCTGAGAGGAATCATCCCGTTAGTAATCTCCCGTAACATATTT 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1055 ttacaanaatccctatagatgtgtaataaacccttaactatagttgcatttctgt 1114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 415 TTATTAATAATCCAAATCTGTGTATTACAAAGCTTGCCAATGTCGCATTAACCCCTT 474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1115 ctccctatataccctgcagagctctctgcagctgtctataactatttaagccacca 1174
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 475 GGCACCTGGCTATTGTCAGAGAGACTTACGGCAGCTGTGACGCTTCGCAATGGAACCA 534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1175 gtatagagattccacacttggtagaaccctggttacagtgtagaacaacagctggatt 1234
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 535 GTACAGAGAAGTTCCACCCCTGTTAGATAGATGCGCGCCAGAGACACTTGTGCT 594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1235 actaagttttctctgcgtatggtcctacagcctcgtcctaaccgataagaag 1294
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 595 CCTCAGCTTCTTTTGGCTGTGCTGACAGCTGTTTACAGTCTCATACCCCAATGAGGAG 654
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1295 gtcaagaagatatgtttcctcaaatgcttatcagaagcttactgaataatgaana 1354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 655 ATCTTACAGATACAAAGCTCACTCACTGAGCTTACAAAGCGTTCAACAAAGAGAGA 714
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1355 ctcttggaatgagagaagaagtttgagaagaatgataatctcctctgcatataag 1414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 715 TGCCGTGGGTAGAGCATGATGCTGAGAGATGAGATTATGTCCTGGGGATGTGGG 774
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1415 ccttggttactcttccctcctgcagcactctatcccttcagtgaagcaatgtttaa 1474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 775 GCTGGCCATCTTGGCTCTTGGCTGTGACATCTATTCACCTGTGAGCGACTCTTTAAC 834
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1475 ctgagaagaattcagttattatcagcttacaactgatatcgcctgcatataagac 1534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 835 CTGGAGAGAAATTTCACTATATTCAGAGCAAACTAGGAATGTCTCTCTCTCTGGGAC 894
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1535 ttccatgttttaattatgtagaagaacagagcttctgagaagaagtagtactacattta 1594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 895 AGTACACGCTTTGGTTTGGCTGGAATAAATGGGTAGATGTCAATCATTTGTATGTA 954
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1595 taacacccaacttctgtctcctgttctgctccccaatgtcaattctgggtaagat 1654
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 955 CATGCTCCGACTTTATGATAGTGTGTTCTCTCAACTCTTGTCTCATATGTAAAT 1014
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY 1655 tatttattccttcacgtatagaagccgaagctaaacgaattaaagaagctgggaaaa 1714
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1015 CCCCCGTGCTCCCTGCTGCTGTGAGAGAGAAAGATACGAGAGATTAGATGTGTTGGGAAGA 1074
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
BB603988
LOCUS
DEFINITION BB603988 Riken full-length enriched, 15 days embryo head mus
musculus cDNA clone D930007L06 5', mRNA sequence.
ACCESSION BB603988
VERSION BB603988.2 GI:16450725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
1 (bases 1 to 680)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

```

```

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Dec 5, 2000 this sequence version replaced gi:11555390.
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/

```

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagl,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome-gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1..680
/organism="Mus musculus"
/strain="C57BL/6J"
/db\_xref="taxon:10090"
/clone="D930007L06"
/clone\_lib="RIKEN full-length enriched, 15 days embryo

FEATURES
SOURCE

1..680
/organism="Mus musculus"
/strain="C57BL/6J"
/db\_xref="taxon:10090"
/clone="D930007L06"
/clone\_lib="RIKEN full-length enriched, 15 days embryo



```
Db 103 CCGTGTGGGACCTGAGGCATCTGCTGTGGCCAAAGTCTCATTTGATGTAGCAACACAA 44
Oy 673 tgaggataaaccaatgacccagatccaa 700
Db 43 TGAGAGTAACCAAGTACCCAGATCCAA 16
```

Search completed: May 9, 2002, 00:02:57  
Job time: 13167 sec

**THIS PAGE BLANK (USPTO)**